



TUGAS AKHIR - SS141501

***CLASSIFICATION AND REGRESSION TREE UNTUK
PENGKLASIFIKASIAN RUMAH TANGGA DENGAN
MALARIA DI PROVINSI PAPUA BARAT DENGAN PRA-
PEMROSESAN *SYNTHETIC MINORITY*
OVERSAMPLING TECHNIQUE***

Ayu Widya Ningrum
NRP 1311 100 121

Dosen Pembimbing
Dr. Drs. I Nyoman Latra, MS.

Co. Dosen Pembimbing
Dr. Mochamad Setyo Pramono, S.Si, M.Si.

Program Studi S1 Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2015



TUGAS AKHIR - SS141501

***CLASSIFICATION AND REGRESSION TREE UNTUK
PENGKLASIFIKASIAN RUMAH TANGGA DENGAN
MALARIA DI PROVINSI PAPUA BARAT DENGAN PRA-
PEMROSESAN *SYNTHETIC MINORITY*
OVERSAMPLING TECHNIQUE***

Ayu Widya Ningrum
NRP 1311 100 121

Dosen Pembimbing
Dr. Drs. I Nyoman Latra, MS.

Co. Dosen Pembimbing
Dr. Mochamad Setyo Pramono, S.Si, M.Si.

Program Studi S1 Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2015



FINAL PROJECT - SS141501

***CLASSIFICATION AND REGRESSION TREE FOR
CLASSIFYING HOUSEHOLDS WITH MALARIA IN WEST
PAPUA PROVINCE WITH PRE-PROCESSING
SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE***

Ayu Widya Ningrum
NRP 1311 111 121

Main Supervisor
Dr. Drs. I Nyoman Latra, MS.

Co. Supervisor
Dr. Mochamad Setyo Pramono, S.Si, M.Si.

Undergraduate Programme of Statistics
Faculty Of Mathematics And Natural Science
Sepuluh Nopember Institute Of Technology
Surabaya 2015

LEMBAR PENGESAHAN

**CLASSIFICATION AND REGRESSION TREE UNTUK
PENGKLASIFIKASIAN RUMAH TANGGA DENGAN
MALARIA DI PROVINSI PAPUA BARAT DENGAN PRA-
PEMROSESAN SYNTHETIC MINORITY OVERSAMPLING
TECHNIQUE**

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi S-1 Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :

AYU WIDYA NINGRUM
NRP 1311 100 121

Disetujui oleh Pembimbing Tugas Akhir

Dr. Drs. I Nyoman Latra, MS
NIP. 19511130 197901 1 001

Dr. Mochamad Setyo Pramono, S.Si, M.Si ()
NIP. 19711230 200501 1 002

Mengetahui
Ketua Jurusan Statistika FMIPA-ITS


Dr. Muhammad Mashuri, MT
NIP. 19620408 198701 1 001

JURUSAN
STATISTIKA
SURABAYA, JULI 2015

CLASSIFICATION AND REGRESSION TREE UNTUK PENGKLASIFIKASIAN RUMAH TANGGA DENGAN MALARIA DI PROVINSI PAPUA BARAT DENGAN PRA- PEMROSESAN SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE

Nama Mahasiswa : Ayu Widya Ningrum
NRP : 1311 100 121
Jurusan : Statistika FMIPA – ITS
Dosen Pembimbing : Dr. Drs. I Nyoman Latra, MS.
Co. Dosen Pembimbing : Dr. M. Setyo P., S.Si, M.Si.

ABSTRAK

Insiden malaria pada penduduk Indonesia tahun 2013 adalah 1,9 persen menurun dibanding tahun 2007. Sedangkan prevalensi malaria tahun 2013 adalah 6,0 persen. Provinsi Papua Barat merupakan provinsi dengan insiden dan prevalensi tertinggi ketiga di Indonesia, pada tahun 2013. Walaupun demikian, Papua Barat mengalami peningkatan tajam dalam hal jumlah penderita malaria. Maka diperlukan upaya untuk menanggulangi kasus malaria di Papua Barat, salah satunya adalah mengetahui karakteristik penyebaran penyakit malaria di Provinsi Papua Barat. Metode statistika yang sering digunakan untuk mendapatkan karakteristik penyebaran malaria adalah regresi logistik, namun hasil analisisnya hanya sebatas mendapatkan model dan faktor-faktor yang berpengaruh saja, belum memunculkan faktor utama yang menjadi penyebab penyebaran penyakit malaria serta hasil ketepatan klasifikasi. Maka digunakan pendekatan CART dengan pra-pemrosesan SMOTE untuk mendapatkan faktor yang diduga dominan dalam mempengaruhi hasil klasifikasi status rumah tangga terhadap malaria di Papua Barat serta dapat meningkatkan hasil akurasi. Penerapan pendekatan CART menunjukkan bahwa variabel terpenting yang berpengaruh dalam menentukan status rumah tangga terhadap penyakit malaria yaitu pekerjaan kepala keluarga. Keakuratan klasifikasi yang dihasilkan pohon optimal untuk data learning sebesar 65,3 persen dan untuk data testing sebesar 68,7 persen.

Kata Kunci : CART, Malaria, Provinsi Papua Barat, Rumah Tangga, SMOTE

(Halaman ini sengaja dikosongkan)

CLASSIFICATION AND REGRESSION TREE FOR CLASSIFYING HOUSEHOLDS WITH MALARIA IN WEST PAPUA PROVINCE WITH PRE-PROCESSING SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE

Name : Ayu Widya Ningrum
NRP : 1311 100 121
Department : Statistika FMIPA – ITS
Main Supervisor : Dr. Drs. I Nyoman Latra, MS.
Co. Supervisor : Dr. M. Setyo P., S.Si, M.Si.

ABSTRACT

The incidence of malaria in the Indonesian population in 2013 was 1,9% decline compared to 2007. While the prevalence of malaria in 2013 is 6%.. West Papua is a province with the third highest incidence and prevalence in Indonesia, in 2013. However, West Papua experienced a sharp increase in the number of malaria patients. efforts are needed to tackle cases of malaria in West Papua, one of which was determine the characteristics of the spread of malaria. statistical methods are often used to obtain the characteristics of the spread of malaria is logistic regression, but the result of the analysis was limites to getting the model and factors that influence it, have not led to major factor that cause the spread of malaria as well as the result of classification accuracy. Then used CART approach with pre-processing SMOTE to obtain the alleged dominant factor in influencing result of household status classiification against malaria in West Papua and can increase the accuracy result. CART approach application shows that the most important variable in optimal classification tree for classifying status of households with malaria in West Papua Province is head of household work. Classification accuracy produced by the tree optimal for learning data is 65,3 percent while 68,7 percent for testing data.

Key Words : CART, Classification, Households, Malaria, SMOTE, West Papua Province

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penyusunan Tugas Akhir yang berjudul “***Classification and Regression Tree*** untuk Pengklasifikasian Rumah Tangga dengan Malaria di Provinsi Papua Barat dengan Pra-Pemrosesan ***Synthetic Minority Oversampling Technique***” dapat diselesaikan dengan baik dan tepat pada waktunya.

Terselesaikannya Tugas Akhir ini, tentu bukan hanya karena usaha dari penulis, tetapi ada banyak pihak yang sangat berjasa dan membantu proses pengerjaan Tugas Akhir ini. Penulis mengucapkan terima kasih kepada :

1. Bapak Dr. Drs. I Nyoman Latra, MS selaku dosen pembimbing dan Bapak Dr. Mochamad Setyo Pramono, S.Si., M.Si selaku co. dosen pembimbing yang senantiasa memberikan ilmu, perhatian, bimbingan dan pengarahan baik selama menyelesaikan Tugas Akhir ini.
2. Bapak Dr. Sutikno, S.Si, M.Si dan Ibu Shofi Andari, M.Si selaku dosen penguji yang senantiasa memberikan kritik dan saran demi kesempurnaan Tugas Akhir ini.
3. Bapak Dr. Muhammad Mashuri, MT selaku Ketua Jurusan Statistika ITS.
4. Dra. Lucia Aridinanti, MS selaku Kaprodi S1 Jurusan Statistika ITS.
5. Ibu tercinta atas segala doa restu, semangat dan *support* yang selalu diberikan, serta Alm. Bapak yang saya yakin telah mendoakan dari atas sana,
6. Seluruh dosen jurusan Statistika atas segala ilmu yang diberikan dan kesabaran yang dilimpahkan. Serta seluruh staf dan karyawan jurusan Statistika atas kerja keras dan bantuannya selama ini.
7. Rekan-rekan Sosialita : Dilla, Sinta, Nurul, Theta, Ida, Gita, Fila, Ecy, Irma, dan Friska yang senantiasa memberikan doa, semangat, dan *support* yang menguatkan, serta tempat

berbagi suka dan duka selama masa perkuliahan hingga terselesaikannya Tugas Akhir ini.

8. Rekan-rekan Trio Malaria : Nurul Fadhillah dan Sinta Krisadini atas kebersamaan dan kekompakannya selama penulisan Tugas Akhir ini
9. Teman-teman seperjuangan Wisuda 112 ITS dan keluarga besar Sigma 22 atas segala kritik, saran, dan motivasinya selama masa perkuliahan hingga terselesaikannya Tugas Akhir ini.
10. Mbak Riza, Mas Yopie, Putri, Suwarno, Lely, dan Ayuk atas diskusi terkait metode CART dan SMOTE.
11. Selanjutnya kepada semua pihak yang telah membantu dalam penulisan Tugas Akhir ini yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran sangat diharapkan dari semua pihak untuk perbaikan yang membangun. Besar harapan penulis agar informasi sekecil apapun dalam Tugas Akhir ini akan dapat menambah wawasan pengetahuan dan bermanfaat bagi semua pihak.

Surabaya, Juli 2015

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	iii
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah	6
BAB II TINJAUAN PUSTAKA	
2.1 Statistika Deskriptif	7
2.2 Tabel Kontingensi	7
2.3 Uji Independensi	8
2.4 Data Tidak Seimbang (<i>Imbalanced Data</i>)	9
2.5 Strategi Sampling	10
2.6 <i>Synthetic Minority Oversampling Technique</i> (SMOTE)	11
2.7 <i>Classification and Regression Tree</i> (CART)	15
2.7.1 Pembentukan Pohon Klasifikasi	17
2.7.2 Pemangkasan Pohon Klasifikasi	21
2.7.3 Penentuan Pohon Klasifikasi Optimal	23
2.7.4 Ukuran Ketepatan Klasifikasi	25
2.8 Penyebaran Penyakit Malaria di Indonesia	26
2.9 Faktor-Faktor yang Mempengaruhi Penyebaran Penyakit Malaria di Indonesia	28
2.10 Penelitian Sebelumnya	33
BAB III METODOLOGI PENELITIAN	

3.1	Sumber Data	35
3.2	Kerangka Konsep Penelitian.....	35
3.4	Variabel Penelitian.....	36
3.5	Langkah Analisis Data.....	39
BAB IV ANALISIS DAN PEMBAHASAN		
4.1	Karakteristik Rumah Tangga di Provinsi Papua Barat pada Tahun 2013.....	43
4.1.1	Rumah Tangga dengan Malaria.....	43
4.1.2	Wilayah Tempat Tinggal	45
4.1.3	Jenis Kelamin Kepala Keluarga.....	46
4.1.4	Status Kawin Kepala Keluarga	47
4.1.5	Tingkat Pendidikan Kepala Keluarga	48
4.1.6	Pekerjaan Kepala Keluarga.....	49
4.1.7	Status Ekonomi Rumah Tangga	50
4.1.8	Umur Kepala Keluarga dan Jumlah Pemakaian Air Rumah Tangga.....	51
4.1.9	Kebiasaan Mengolah Air Minum Sebelum Dikonsumsi	52
4.1.10	Adanya Layanan Kesehatan Gratis	53
4.1.11	Jenis Sumber Air Utama Rumah Tangga.....	54
4.1.12	Adanya pencegahan Gigitan Nyamuk	55
4.1.13	Lingkungan Tempat Tinggal	55
4.1.14	Kepadatan Hunian.....	56
4.2	Uji Independensi.....	57
4.3	Klasifikasi Penderita Penyakit Malaria di Provinsi Papua Barat dengan Pohon Klasifikasi.....	58
4.3.1	Pembentukan Pohon Klasifikasi Maksimal	61
4.3.2	Pemangkasan Pohon Klasifikasi Maksimal (<i>Pruning</i>).....	63
4.3.3	Pemilihan Pohon Klasifikasi Optimal.....	64
4.3.4	Hasil Ketepatan Klasifikasi Pohon Klasifikasi CART.....	69
BAB V KESIMPULAN DAN SARAN		
5.1	Kesimpulan.....	73
5.2	Saran	74

DAFTAR PUSTAKA.....	75
LAMPIRAN.....	79
BIODATA PENULIS	

(Halaman Ini Sengaja Dikosongkan)

DAFTAR TABEL

	Halaman
Tabel 2.1 Tabel Kontingensi $I \times J$	8
Tabel 2.2 Tabel Probabilitas $I \times J$	8
Tabel 2.3 Data Simulasi	12
Tabel 2.4 Distribusi Data Simulasi	14
Tabel 2.5 Data simulasi setelah menggunakan SMOTE	14
Tabel 2.6 <i>Crosstab</i> Ketepatan Klasifikasi	25
Tabel 3.1 Struktur Data Penelitian	35
Tabel 3.2 Variabel Respon	36
Tabel 3.3 Variabel Penelitian Demografi dan Sosial Ekonomi Responden	37
Tabel 3.4 Variabel Penelitian Faktor Resiko	38
Tabel 4.1 Karakteristik Rumah Tangga Berdasarkan Wilayah Tempat Tinggal	46
Tabel 4.2 Karakteristik Rumah Tangga Berdasarkan Jenis Kelamin Kepala Keluarga	46
Tabel 4.3 Karakteristik Rumah Tangga Berdasarkan Status Kawin Kepala Keluarga	48
Tabel 4.4 Karakteristik Rumah Tangga Berdasarkan Tingkat Pendidikan Kepala Keluarga	49
Tabel 4.5 Karakteristik Rumah Tangga Berdasarkan Pekerjaan Kepala Rumah Tangga	50
Tabel 4.6 Karakteristik Rumah Tangga Berdasarkan Status Ekonomi	51
Tabel 4.7 Karakteristik Rumah Tangga Berdasarkan Umur Kepala Keluarga dan Jumlah Pemakaian Air	52
Tabel 4.8 Karakteristik Rumah Tangga Berdasarkan Pengolahan Air Minum	52
Tabel 4.9 Karakteristik Rumah Tangga Berdasarkan . Adanya Layanan Kesehatan Gratis	53

Tabel 4.10	Karakteristik Rumah Tangga Berdasarkan Jenis Sumber Air Utama.....	54
Tabel 4.11	Karakteristik Rumah Tangga Berdasarkan Adanya Pencegahan Gigitan Nyamuk	55
Tabel 4.12	Karakteristik Rumah Tangga Berdasarkan Lingkungan Tempat Tinggal	56
Tabel 4.13	Karakteristik Rumah Tangga Berdasarkan Kepadatan Hunian	57
Tabel 4.14	Uji Independensi.....	57
Tabel 4.15	Ketepatan Klasifikasi Sebelum dan Sesudah Prapemrosesan SMOTE.....	60
Tabel 4.16	Variabel Penting Pembentukan Pohon Klasifikasi .. Maksimal.....	61
Tabel 4.17	Variabel Penting Pembentukan Pohon Klasifikasi .. Optimal.....	66
Tabel 4.18	Kelas Rumah Tangga pada Masing-masing Simpul Terminal.....	69
Tabel 4.19	Karakteristik Kelas Rumah Tangga Menurut Persentase Kelas Tertinggi Simpul Terminal	69
Tabel 4.20	Klasifikasi Data <i>Learning</i> oleh Pohon Klasifikasi .. Optimal.....	70
Tabel 4.21	Klasifikasi Data <i>Testing</i> oleh Pohon Klasifikasi Optimal.....	70
Tabel 4.22	Perbandingan Ketepatan Klasifikasi Pohon Maksimal dan Pohon Optimal	71

DAFTAR GAMBAR

	Halaman
Gambar 2. 1 Ilustrasi Algoritma <i>SMOTE</i>	12
Gambar 2. 2 Persentase Masing-Masing Kelas.....	13
Gambar 2. 3 Ilustrasi Pohon Klasifikasi	16
Gambar 2.4 <i>Annual Parasite Incidence</i> (API) per 1,000 Penduduk di Indonesia Tahun 2008-2013	28
Gambar 3.1 Kerangka Konsep Penelitian (Pengembangan Faktor Status Kesehatan Blum).....	36
Gambar 3.2 Diagram Alir Analisis Data.....	41
Gambar 4.1 Karakteristik Rumah Tangga Penderita Penyakit Malaria	44
Gambar 4.2 Karakteristik Rumah Tangga per Kelas Pada Tiap Kabupaten/Kota di Provinsi Papua Barat.....	44
Gambar 4.3 Topologi Pohon Klasifikasi Maksimal.....	63
Gambar 4.4 Plot <i>Relative Cost</i> dan Banyaknya Simpul Terminal	64
Gambar 4.5 Topologi Pohon Klasifikasi Optimal.....	64
Gambar 4.6 Potongan Struktur Pohon Klasifikasi Optimal	68

(Halaman Ini Sengaja Dikosongkan)

BIOGRAFI PENULIS



Ayu Widya Ningrum atau yang lebih dikenal dengan sapaan Ayu atau Nduty terlahir di Kota Pahlawan pada tanggal 18 Agustus 1992, Ayu merupakan anak ketiga dari tiga bersaudara dengan kakak perempuan bernama Yukanti Sriyati Ningsih dan kakak laki-laki bernama Desembry Yanto. Putri dari pasangan Bapak Triyono dan Ibu Ningsih ini menempuh jenjang pendidikan formal mulai dari TK Hang Tuah XI, SDN Kebonsari I/414, SMPN 22 Surabaya, SMAN 16 Surabaya dan pada tahun 2011 ia diterima menjadi mahasiswa Jurusan Statistika ITS. Selain menjalani aktifitas akademik, Ayu juga terlibat dalam beberapa kegiatan ekstrakurikuler, kepanitiaan dan kegiatan berorganisasi. Ayu juga memiliki pengalaman menjadi anggota tim Program Kreatifitas Mahasiswa Bidang Penelitian yang didanai Dikti dan menjalani Kerja Praktek di Perusahaan Gas Negara. Ayu pernah magang di Bank Indonesia Surabaya untuk mengisi waktu liburan semester. Ayu sangat menggemari dunia kuliner, sehingga memasak adalah hal yang paling disukainya. Selain menggemari memasak, Ayu juga suka membagi ilmu dan wawasan kepada orang lain, ia mengisi waktu luangnya dengan memberikan bimbingan belajar. Apabila pembaca memiliki saran, kritik, atau ingin berdiskusi dengan penulis tentang Tugas Akhir atau wawasan lain, silahkan kirim email ke ayuwidyaningrum121@gmail.com.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Malaria merupakan penyakit menular yang menjadi perhatian global, termasuk di Indonesia. Penyakit ini masih merupakan masalah kesehatan masyarakat karena sering menimbulkan kejadian luar biasa (KLB), berdampak luas terhadap kualitas hidup dan ekonomi, serta dapat mengakibatkan kematian. Penyakit ini dapat bersifat akut, laten atau kronis (Kementerian Kesehatan RI, 2013). Penularan parasit *Plasmodium sp.* kepada manusia adalah melalui nyamuk *Anopheles* betina yaitu ketika menggigit manusia, nyamuk tersebut menyuntikkan parasit ke dalam aliran darah menuju hati kemudian melipatgandakan diri (Kementerian Kesehatan RI, 2014).

Insiden malaria pada penduduk Indonesia tahun 2013 adalah 1,9 persen menurun dibanding tahun 2007 (2,9%). Sedangkan prevalensi malaria tahun 2013 adalah 6,0 persen. Meskipun demikian, Papua Barat mengalami peningkatan tajam dalam hal jumlah penderita malaria dibandingkan dengan provinsi lain di Indonesia. Lima Provinsi dengan insiden dan prevalensi tertinggi adalah Papua (9,8% dan 28,6%), Nusa Tenggara Timur (6,8% dan 23,3%), Papua Barat (6,7% dan 19,4%), Sulawesi Tengah (5,1% dan 12,5%), dan Maluku (3,8% dan 10,7%). Sebanyak 15 provinsi dari 33 provinsi di Indonesia mempunyai prevalensi malaria diatas angka nasional, sebagian besar berada di Indonesia Timur (Kementerian Kesehatan RI, 2013).

Provinsi Papua Barat merupakan provinsi dengan insiden dan prevalensi tertinggi ketiga di Indonesia setelah Provinsi Papua dan Nusa Tenggara Timur. Sedangkan *annual parasite incidence* (API) di Papua Barat merupakan API tertinggi kedua yaitu sebesar 38,44 per 1000 penduduk setelah Papua. Namun Provinsi Papua Barat mengalami peningkatan tajam dalam hal jumlah penderita malaria pada tahun 2013. Sehingga diperlukan

upaya untuk menanggulangi kasus malaria di Provinsi Papua Barat agar angka kejadian malaria di Papua Barat dapat ditekan. Berbagai upaya perlu dilakukan untuk menanggulangi kasus malaria di Provinsi Papua Barat. Salah satu upaya yang perlu dilakukan adalah mengetahui karakteristik penyebaran penyakit malaria di Provinsi Papua Barat sebelum melakukan tindakan yang lebih lanjut. Beberapa penerapan metode statistika telah banyak digunakan untuk mengetahui karakteristik penderita malaria.

Penelitian tentang penyakit malaria di kawasan Indonesia Timur telah banyak dilakukan, diantaranya adalah penelitian yang dilakukan oleh Ekayani (2011), Lestari (2014), dan Susilowati (2014). Ekayani dan Lestari menggunakan regresi logistik biner, sedangkan Susilowati menggunakan regresi *ordinary least square* (CART) dan *robust*. Berdasarkan ketiga penelitian tersebut didapatkan faktor-faktor yang mempengaruhi penyakit malaria antara lain adalah pekerjaan, saluran pembuangan limbah, tempat penampungan limbah, penggunaan kelambu, sarana penampungan air minum, pemanfaatan posyandu dan pemanfaatan POD (Pos Obat Desa)/WOD (Warung Obat Desa), akses air bersih, rumah panggung, atap ijuk/rumbia, atap seng, dan lantai semen plesteran retak. Berdasarkan penelitian-penelitian tersebut, masih pada tahap mendapatkan faktor-faktor yang berpengaruh, namun belum memunculkan model klasifikasi dan faktor utama yang menjadi pangkal permasalahan penyakit malaria serta belum memunculkan tingkat akurasi dari suatu pengklasifikasian.

Metode yang umum digunakan dalam proses klasifikasi adalah analisis diskriminan dan regresi logistik multivariat. Metode ini memiliki keterbatasan dalam hal pemenuhan asumsi dan kesederhanaan interpretasi. Analisis diskriminan mensyaratkan terpenuhinya asumsi kenormalan secara multivariat untuk variabel prediktor dan varians homogen untuk setiap kelas dalam variabel respon. Sedangkan, regresi logistik multivariat dalam penelitian Purwanto (2009) memiliki ketepatan klasifikasi yang rendah dan hasil klasifikasi yang relatif sulit untuk diinter-

pretasikan. Salah satu metode pengklasifikasian yang bersifat nonparametrik dan mampu mengatasi keterbatasan kemampuan klasifikasi dari metode yang telah umum digunakan adalah metode *classification and regression tree* (CART). Dalam metode ini, proses klasifikasi dilakukan melalui penyekatan rekursif biner. Metode CART dapat menyeleksi variabel-variabel prediktor yang paling penting dalam menentukan hasil klasifikasi variabel respon. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. CART mempunyai beberapa kelebihan antara lain mampu bekerja pada dimensi data yang besar dan struktur data yang kompleks, tidak terikat oleh asumsi kenormalan maupun variansi homogen, dapat mengetahui interaksi antar variabel prediktor dan hasil klasifikasi yang diperoleh lebih mudah dipahami serta diinterpretasikan karena struktur datanya dapat dilihat secara visual (Lewis dan Roger, 2000). Beberapa penelitian dengan menerapkan metode CART pernah dilakukan oleh Irawan (2014) tentang klasifikasi status HIV/AIDS di LSM Orbit Surabaya menghasilkan kesimpulan bahwa ketepatan klasifikasi status HIV/AIDS sebesar 65% dan hasil validasi pada data tes sebesar 73,68%. Penelitian lain dilakukan oleh Seftiana (2014) tentang klasifikasi rumah tangga sangat miskin di kabupaten Jombang dengan metode RF-CART menghasilkan tingkat akurasi sebesar 65,5% untuk data *learning* dan 62,8% untuk data *testing*. Berdasarkan penelitian-penelitian dengan menggunakan metode CART, tingkat akurasi yang diperoleh masih relatif rendah sehingga diperlukan upaya untuk meningkatkan nilai akurasi dalam pengklasifikasian.

Suatu pengklasifikasian memiliki tingkat akurasi rendah dikarenakan jumlah data masing-masing kelas tidak seimbang, dimana terdapat satu kelas yang memiliki jumlah data yang kecil bila dibandingkan dengan kelas lainnya (Chawla, Bowyer, Hall, dan Kegelmeyer, 2002). Salah satu metode yang mampu mengatasi masalah ketidak seimbangan pada data adalah metode *synthetic minority oversampling technique* (SMOTE). Metode

SMOTE ini merupakan salah satu metode *oversampling*. Pendekatan *oversampling* bekerja dengan *synthetic data* yaitu data replikasi dari data minor. Pendekatan *oversampling* dilakukan dengan cara mereplikasi data minor sehingga tidak mengurangi banyak informasi seperti yang dilakukan oleh pendekatan *undersampling*. Peran metode SMOTE disini adalah untuk menyeimbangkan data tersebut dengan cara menduplikasi kelas *minority*. Penelitian yang menerapkan metode SMOTE pernah dilakukan oleh Trapsilasiwi (2014) mengenai kasus kanker payudara dan kanker serviks dan menghasilkan kesimpulan bahwa pada klasifikasi kanker payudara menghasilkan tingkat akurasi sebesar 89,2% sebelum menggunakan SMOTE dan 96,8% sesudah menggunakan SMOTE. Sedangkan pada klasifikasi kanker serviks, menghasilkan ketepatan akurasi sebesar 40,4% sebelum menggunakan SMOTE dan 59,3% sesudah menggunakan SMOTE. Berdasarkan penelitian tersebut, dapat disimpulkan bahwa dengan menggunakan metode SMOTE, tingkat akurasi yang diperoleh lebih tinggi bila dibandingkan tanpa SMOTE. Sehingga sebelum melakukan analisis dengan metode ini, perlu dilakukan tahap pra-pemrosesan menggunakan metode SMOTE agar hasil ketepatan klasifikasinya lebih tinggi.

Pada penelitian ini akan digunakan metode *classification and regression tree* (CART) untuk memodelkan klasifikasi status rumah tangga dengan malaria dan mendapatkan faktor yang paling dominan mempengaruhi hasil klasifikasi status rumah tangga dengan malaria di Papua Barat. Sebelum memodelkan klasifikasi penderita malaria di Provinsi Papua Barat, perlu menerapkan metode *synthetic minority oversampling technique* (SMOTE) sebagai tahap pra-pemrosesan, sehingga kedepannya dapat meningkatkan nilai akurasi dari hasil pengklasifikasian. Penelitian ini diharapkan dapat membantu Pemerintah Daerah Provinsi Papua Barat dalam menekan angka penderita malaria.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan bahwa Provinsi Papua Barat merupakan provinsi yang mengalami

peningkatan tajam dalam hal jumlah penderita malaria dibandingkan dengan provinsi lain di Indonesia. Sehingga diperlukan upaya untuk menanggulangi kasus malaria di Provinsi Papua Barat agar angka kejadian malaria di Papua Barat dapat ditekan. Salah satu upaya yang dilakukan adalah melakukan analisis klasifikasi status rumah tangga terhadap penyakit malaria di Provinsi Papua Barat. Sehingga rumusan masalah penelitian ini adalah bagaimana hasil klasifikasi rumah tangga terhadap penyakit malaria di Provinsi Papua Barat berdasarkan faktor-faktor yang mempengaruhi menggunakan pendekatan *classification and regression tree* (CART) dengan pra-pemrosesan *synthetic minority oversampling technique* (SMOTE)?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini secara umum yaitu mendapatkan klasifikasi rumah tangga dengan malaria di Provinsi Papua Barat berdasarkan faktor-faktor yang mempengaruhi menggunakan pendekatan *classification and regression tree* (CART) dengan pra-pemrosesan *synthetic minority oversampling technique* (SMOTE). Sedangkan tujuan khusus dari penelitian ini adalah sebagai berikut.

- a. Mendapatkan model klasifikasi status rumah tangga dengan malaria di Provinsi Papua Barat.
- b. Mengetahui faktor atau variabel yang paling dominan mempengaruhi penentuan hasil klasifikasi rumah tangga dengan malaria di Provinsi Papua Barat.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dalam penelitian ini antara lain adalah sebagai berikut.

1. Memberikan informasi ke Pemerintah Daerah Provinsi Papua Barat mengenai klasifikasi rumah tangga dengan malaria di Provinsi Papua Barat beserta faktor yang berpengaruh menggunakan metode *classification and regression tree* (CART), guna membantu *policy* atau kebijakan Pemerintah

Daerah Provinsi Papua Barat untuk menurunkan angka penderita malaria.

2. Bagi peneliti sebagai sarana mengaplikasikan metode *classification and regression tree* (CART) dalam mengatasi masalah epidemiologi malaria supaya bermanfaat secara nyata untuk masyarakat dalam dunia kesehatan.

1.5 Batasan Masalah

Batasan masalah penelitian adalah bahwa data yang digunakan hanya data jumlah kejadian penyakit malaria di Provinsi Papua Barat pada tahun 2013 yang diambil dari Laporan Riset Kesehatan Dasar (Riskesdas) 2013. Variabel respon pada penelitian ini bersifat kategorik, sehingga digunakan metode klasifikasi pohon (*classification tree*). Pemilah terbaik yang digunakan adalah indeks Gini dan indeks Twoing.

BAB II

TINJAUAN PUSTAKA

2.1 Statistika Deskriptif

Statistika deskriptif adalah suatu metode analisis statistika yang berkaitan dengan pengumpulan dan penyajian data yang biasanya disajikan dalam bentuk tabel, grafik, nilai pemusatan dan atau nilai penyebaran. Statistika deskriptif sering disebut sebagai statistika deduktif karena kesimpulan analisis yang diambil dengan metode ini adalah kesimpulan yang bersifat deduktif bukan induktif atau inferensi (Walpole, 1995).

2.2 Tabel Kontingensi (*Contingency Table*)

Tabel kontingensi atau yang sering disebut tabulasi silang (*cross-tabulation*) adalah tabel yang berisi data jumlah atau frekuensi dari beberapa kategori (klasifikasi). Tabel kontingensi merupakan metode yang umumnya digunakan untuk meringkas data kategorikal. Umumnya, tabel kontingensi digunakan untuk melihat hubungan antara variabel pada baris dan variabel pada kolom, kadangkala tabel kontingensi digunakan untuk mencari informasi lebih lanjut untuk melihat besarnya hubungan antar variabel tersebut. Data yang digunakan dapat berasal dari beberapa kerangka sampling yang berbeda-beda, dan interpretasi dari hipotesis mengenai tidak adanya hubungan tergantung dari kerangka yang dibangun (Stokes, Davis, dan Koch, 2000).

Misalkan terdapat variabel X dan variabel Y yang merupakan variabel kategorikal, dengan X memiliki kategori sebanyak I dan Y memiliki kategori sebanyak J . Klasifikasi untuk setiap subjek dari kedua variabel memiliki kombinasi sebanyak IJ . Tabel kontingensi akan memiliki baris sebanyak I untuk kategori pada variabel X dan memiliki kolom sebanyak J untuk kategori pada variabel Y .

Tabel 2.1. Tabel kontingensi $I \times J$

Variabel X	Variabel Y				Total
	Y_1	Y_2	\dots	Y_J	
X_1	n_{11}	n_{12}	\dots	n_{1J}	$n_{1.}$
X_2	n_{21}	n_{22}	\dots	n_{2J}	$n_{2.}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
X_I	n_{I1}	n_{I2}	\dots	n_{IJ}	$n_{I.}$
Total	$n_{.1}$	$n_{.2}$	\dots	$n_{.J}$	$n_{..}$

Tiap sel pada tabel kontingensi yang ditunjukkan pada Tabel 1 menunjukkan frekuensi atau banyaknya kejadian dari kategori (i,j) , dengan $i=1,2,\dots,I$ dan $j=1,2,\dots,J$.

Untuk masing-masing kejadian, dapat dihitung kemungkinan atau probabilitas dari masing-masing kejadian dalam sel.

Tabel 2.2. Tabel probabilitas kontingensi $I \times J$

Variabel X	Variabel Y				Total
	Y_1	Y_2	\dots	Y_J	
X_1	P_{11}	P_{12}	\dots	P_{1J}	$P_{1.}$
X_2	P_{21}	P_{22}	\dots	P_{2J}	$P_{2.}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
X_I	P_{I1}	P_{I2}	\dots	P_{IJ}	$P_{I.}$
Total	$P_{.1}$	$P_{.2}$	\dots	$P_{.J}$	$P=1$

2.3 Uji Independensi

Dalam menjelaskan hubungan antara variabel respon dan variabel eksplanatori atau prediktor, maka diperlukan adanya keterkaitan antar variabel respon dengan variabel prediktor. Uji independensi digunakan untuk mengetahui adanya keterkaitan antar variabel (Agresti, 2002). Uji independensi merupakan pengujian non-parametrik untuk hipotesis awal yang menyatakan bahwa tidak terdapat hubungan antara variabel respon dengan variabel prediktor. Hipotesis untuk pengujian independensi dapat dituliskan sebagai berikut:

H_0 : Tidak terdapat hubungan antara variabel respon dengan variabel prediktor

H_1 : Terdapat hubungan antara variabel respon dengan variabel prediktor

Statistik uji :

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (2.1)$$

dengan,

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n_{..}} \quad (2.2)$$

di mana :

χ^2 = nilai statistik uji independensi

n_{ij} = frekuensi pengamatan pada kolom ke- i baris ke- j

E_{ij} = nilai ekspektasi pengamatan pada kolom ke- i baris ke- j

$n_{i.}$ = frekuensi pengamatan pada kolom ke- i

$n_{.j}$ = frekuensi pengamatan pada baris ke- j

i = banyak kategori pada variabel prediktor ($i = 1, 2, \dots, I$)

j = banyak kategori pada variabel respon ($j = 1, 2, \dots, J$)

Hipotesis awal ditolak jika nilai χ^2 yang didapatkan berdasarkan perhitungan memiliki nilai lebih dari nilai titik kritis $\chi^2_{(I-1)(J-1)}$ pada batas kesalahan yang ditentukan sebesar α (Daniel, 1989).

Uji independensi pada penelitian ini digunakan untuk melihat adanya keterkaitan antara penyakit malaria yang menyerang rumah tangga di Provinsi Papua Barat tahun 2013 dengan faktor-faktor yang dapat menyebabkan rumah tangga terjangkit penyakit malaria.

2.4 Data Tidak Seimbang (*Imbalanced Data*)

Data tidak seimbang merupakan kondisi dimana terdapat satu kelas yang memiliki jumlah *instance* yang kecil bila dibandingkan dengan kelas lainnya. Kelas yang memiliki jumlah *instance* yang kecil disebut *minority* dan kelas yang memiliki

jumlah *instance* besar disebut *majority* (Chawla, Bowyer, Hall, dan Kegelmeyer, 2002). Permasalahannya hal yang ingin diamati ialah kelas *minority* sehingga sering terjadi kesalahan klasifikasi pada kelas *minority*. Ketidakseimbangan data ini dapat diatasi dengan beberapa cara, di antaranya dengan pengambilan sampel pada tiap kelas dan strategi *sampling* seperti *oversampling* dan *undersampling*.

2.5 Strategi Sampling

Sampling merupakan bagian dari ilmu statistik yang memfokuskan penelitian terhadap pemilihan data yang dihasilkan dari satu kumpulan populasi data. Metode sampling atau yang lebih dikenal dengan *resample* adalah metode umum yang digunakan dalam menyelesaikan permasalahan *imbalance* data. Dengan adanya penerapan sampling pada data yang *imbalance*, tingkat *imbalance* semakin kecil dan klasifikasi dapat dilakukan dengan tepat. Strategi *sampling* merupakan salah satu teknik yang populer dalam mengatasi ketidakseimbangan data. Strategi *sampling* akan mendistribusikan data pada 2 kelas mendekati sama. Teknik strategi sampling di antaranya *oversampling* kelas *minority* atau *undersampling* kelas *majority* (Chawla dkk, 2002).

Strategi *undersampling* dilakukan pada kelas *majority* sehingga jumlah *instance* kelas *majority* sama dengan jumlah kelas *minority*. Strategi ini dapat dilakukan dengan memilih secara acak kelas *majority*. Strategi *undersampling* dengan pemilihan acak dapat menyebabkan pemilihan *instance* tidak mewakili populasi. Oleh karena itu, dilakukan *cluster* pada data *majority* sebelum dilakukan pemilihan data.

Strategi *oversampling* dilakukan pada data kelas *minority* sehingga jumlah kelas *minority* mendekati jumlah kelas *majority*. Strategi ini dapat dilakukan dengan menduplikasi kelas *minority*. Strategi *oversampling* dengan duplikasi memiliki beberapa *instance* yang sama sehingga tidak memiliki variasi data. Oleh karena itu, strategi *oversampling* juga dilakukan dengan

pembangkitan data acak untuk masing-masing atribut independen. Hal ini menghasilkan *instance* dengan kombinasi nilai atribut berbeda dengan data aslinya.

2.6 Synthetic Minority Oversampling Technique (SMOTE)

Synthetic Minority Oversampling Technique (SMOTE) merupakan salah satu metode *oversampling* yaitu teknik pengambilan sampel untuk meningkatkan jumlah data pada kelas positif dengan cara mereplikasi jumlah data pada kelas positif secara acak sehingga jumlahnya sama dengan data pada kelas negatif. Algoritma SMOTE pertama kali ditemukan oleh Chawla (2002). Pendekatan ini bekerja dengan membuat “*synthetic*” data, yaitu data replikasi dari data minor. Metode SMOTE bekerja dengan mencari *k nearest neighbors* (ketetanggaan data). Teknik ini termasuk dalam kelompok klasifikasi non parametrik. Mirip dengan *clustering*, teknik ini sangat sederhana dan mudah untuk diimplementasikan. Teknik ini bekerja dengan mengelompokkan data berdasarkan tetangga terdekat. Tetangga terdekat dipilih berdasarkan jarak *euclidean* antara kedua data. Misalkan diberikan dua data dengan p dimensi yaitu $\mathbf{x}^T = [x_1, x_2, \dots, x_p]$ dan $\mathbf{y}^T = [y_1, y_2, \dots, y_p]$ maka jarak *euclidean* $d(\mathbf{x}, \mathbf{y})$ antara kedua vektor data adalah sebagai berikut,

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \quad (2.3)$$

sedangkan *synthetic* data dilakukan dengan menggunakan Persamaan (2.4)

$$x_{syn} = x_i + (x_{knn} - x_i) \times \beta, \quad i=1, 2, \dots, n \quad (2.4)$$

dengan,

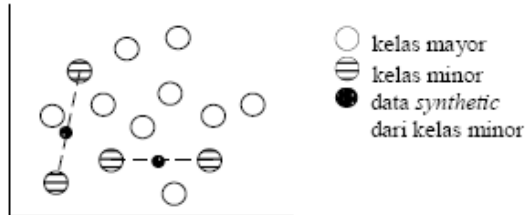
x_{syn} = data hasil replikasi

x_i = data yang akan direplikasi

x_{knn} = data yang memiliki jarak terdekat dari data yang akan direplikasi

β = bilangan random antara 0 sampai 1

Ilustrasi distribusi data setelah diterapkan metode SMOTE dapat dilihat pada Gambar 2.1.



Gambar 2. 1 Ilustrasi Algoritma *SMOTE*

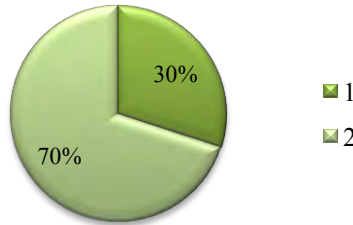
(Sumber : Chawla, Bowyer, Hall, dan Kegelmeyer, 2002)

Sebagai contoh mengenai ilustrasi algoritma SMOTE maka dilakukan simulasi yang ditunjukkan pada Tabel 2.1, misalkan variabel prediktor yang digunakan yaitu jumlah pemakaian air (X_1) dan umur kepala keluarga (X_2) serta variabel Y terdiri dari 2 kelas yaitu kelas 1 (terjangkit malaria) berjumlah 3 dan kelas 2 (tidak terjangkit malaria) berjumlah 7.

Tabel 2. 3 Data Simulasi

Data ke-	Jumlah Pemakaian Air	Umur KK	Y
1	100	32	1
2	200	30	1
3	300	33	1
4	200	58	2
5	100	65	2
6	200	72	2
7	100	56	2
8	300	83	2
9	150	52	2
10	200	51	2

Berdasarkan Tabel 2.3, dapat dilihat persentase untuk masing-masing kelas yang ditunjukkan oleh Gambar 2.2



Gambar 2. 2 Persentase Masing-Masing Kelas Pada Data Simulasi

Gambar 2.2 memberikan informasi bahwa persentase untuk masing-masing kelas tidak seimbang. Terdapat perbedaan jumlah anggota yang mencolok antara kelas 1 dan 2. Hal ini yang disebut *imbalanced data*.

Data yang tidak seimbang tersebut kemudian diolah dengan menggunakan algoritma SMOTE. Tahapan yang dilakukan pada algoritma SMOTE adalah sebagai berikut.

1. Setiap data pada kelas minor yang akan direplikasi mencari tetangga terdekat (x_{knn}) dengan menggunakan jarak *euclidean*. Misalkan data yang akan direplikasi (x_i) pada kelas 1 yaitu data simulasi dengan koordinat (100,32). Algoritma *nearest neighbor* akan menghitung jarak antara koordinat (100,32) dengan (200,30) dan koordinat (100,32) dengan (300,33). Menghitung jarak *euclidean* dengan menggunakan persamaan 2.3,

$$d\left(\begin{bmatrix} 100 \\ 32 \end{bmatrix}, \begin{bmatrix} 200 \\ 30 \end{bmatrix}\right) = \sqrt{(100 - 200)^2 + (32 - 30)^2} = 100,02$$

$$d\left(\begin{bmatrix} 100 \\ 32 \end{bmatrix}, \begin{bmatrix} 300 \\ 33 \end{bmatrix}\right) = \sqrt{(100 - 300)^2 + (32 - 33)^2} = 200$$

Berdasarkan perhitungan tersebut dapat diketahui bahwa data dengan koordinat (200,30) memiliki jarak terpendek terhadap data dengan koordinat (100,32). Sehingga (x_{knn}) dari x_i adalah data dengan koordinat (200,30).

2. Menghitung *synthetic data* dengan menggunakan rumus pada persamaan 2.4, perhitungan data sintesis (data hasil replikasi) pada kelas 1 adalah sebagai berikut,

$$\begin{aligned} x_{syn} &= [100, 32] + ([200, 30] - [100, 32]) \times 0,2 = [100, 32] + [20; -0,4] \\ &= [120; 31,6] \end{aligned}$$

Data sintesis yang dihasilkan adalah data dengan koordinat (120;31,6).

Distribusi data simulasi dengan menggunakan SMOTE ditunjukkan pada Tabel 2.4.

Tabel 2. 4 Distribusi Data Simulasi

Kelas Mayor	Kelas Minor	Replikasi	Kelas Mayor	Kelas Minor Baru
7 (70%)	3 (30%)	1 kali	7 (54%)	6 (46%)

Replikasi pada algoritma SMOTE tentunya meningkatkan jumlah data simulasi yang awalnya berjumlah 10 menjadi 13. Pada Tabel 2.2 diketahui bahwa data simulasi telah seimbang, hal ini dapat dilihat dari persentase masing-masing kelas. Data simulasi setelah menggunakan SMOTE ditunjukkan pada Tabel 2.5.

Tabel 2. 5 Data simulasi setelah menggunakan SMOTE

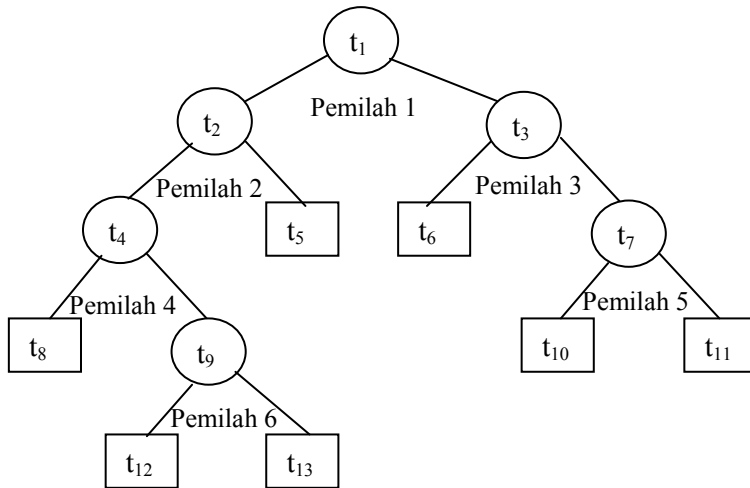
Data ke-	X ₁	X ₂	Y	Data ke-	X ₁	X ₂	Y
1	100	32	1	8	300	83	2
2	200	30	1	9	150	52	2
3	300	33	1	10	200	51	2
4	200	58	2	11*	120	31,6	1
5	100	65	2	12*	180	34	1
6	200	72	2	13*	280	32,4	1
7	100	56	2				

* : *synthetic data*

2.7 *Classification and Regression Tree (CART)*

Classification and Regression Tree (CART) merupakan salah satu metode *machine learning* dimana metode eksplorasi data dilakukan dengan teknik pohon keputusan (*decision tree*). Metode CART merupakan teknik klasifikasi dengan menggunakan algoritma penyekatan rekursif secara biner (*binary recursive partitioning*). Istilah “*binary*” diartikan sebagai pemilahan yang dilakukan pada sekelompok data yang terkumpul dalam suatu ruang yang disebut simpul/*node* menjadi dua kelompok yang disebut sebagai simpul anak (*child nodes*). Istilah “*recursive*” menunjukkan bahwa prosedur penyekatan secara biner dilakukan secara berulang-ulang. Setiap simpul anak yang diperoleh dari penyekatan simpul awal kemudian dapat dipilah kembali menjadi dua simpul anak lagi, dan begitu seterusnya sampai memenuhi kriteria tertentu. Sedangkan, istilah “*partitioning*” memiliki arti bahwa proses klasifikasi dilakukan dengan cara memilah suatu kumpulan data menjadi beberapa bagian atau partisi (Lewis dan Roger, 2000).

Ilustrasi pohon klasifikasi ditunjukkan pada Gambar 2.3. Simpul awal yang merupakan variabel terpenting dalam menduga kelas amatan disebut sebagai *parent node* dengan notasi t_1 , simpul dalam atau *internal nodes* dinotasikan dengan t_2, t_3, t_4, t_7 dan t_9 , serta simpul akhir yang disebut sebagai *terminal nodes* dinotasikan dengan $t_5, t_6, t_8, t_{10}, t_{11}, t_{12}$ dan t_{13} dimana setelahnya tidak ada lagi pemilahan. Setiap simpul berada pada kedalaman (*depth*) tertentu dimana t_1 berada pada kedalaman 1, t_2 dan t_3 berada pada kedalaman 2, dan begitu seterusnya hingga simpul t_{12} dan t_{13} yang berada pada kedalaman 5.



Gambar 2. 3 Ilustrasi Pohon Klasifikasi

(Sumber : Breiman, Friedman, Olshen, dan Stone 1993)

Menurut Breiman, Friedman, Olshen, dan Stone (1993), CART akan menghasilkan pohon klasifikasi jika variabel respon mempunyai skala kategorik dan akan menghasilkan pohon regresi jika variabel respon berupa data kontinu. Metode pengklasifikasian CART memiliki beberapa kelebihan sebagai berikut.

- a. Metode CART bersifat nonparametrik sehingga tidak memerlukan asumsi distribusi variabel prediktor yang perlu dipenuhi (seperti asumsi multivariat normal dan varians homogen).
- b. CART mampu mempertimbangkan interaksi antar variabel.
- c. Metode CART memudahkan dalam hal eksplorasi dan pengambilan keputusan pada struktur data yang kompleks dan multi variabel karena struktur data dapat dilihat secara visual.
- d. Hasil klasifikasi akhir berbentuk sederhana dan dapat mengklasifikasikan data baru secara lebih efisien serta

mudah diinterpretasikan terutama bagi non statistisi (Lewis dan Roger, 2000).

Disamping beberapa kelebihan yang dimiliki, metode CART juga memiliki kelemahan yaitu pohon yang terbentuk tidak stabil artinya ketika terjadi sedikit perubahan pada data *learning* maka hasil prediksi pohon yang diperoleh dapat mengalami perbedaan yang cukup besar (Sutton, 2005). Algoritma CART secara umum melalui tiga tahapan yaitu pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi, dan penentuan pohon klasifikasi optimum.

2.7.1 Pembentukan Pohon Klasifikasi

Proses pembentukan pohon klasifikasi harus memperhatikan tiga hal penting yaitu memilih pemilah, memutuskan apakah simpul perlu dipilah lagi atau sudah dijadikan simpul terminal, dan memberi label kelas pada setiap simpul terminal yang terbentuk (Breiman dkk, 1993). Dalam proses pembentukan pohon klasifikasi membutuhkan data *learning*, sehingga perlu dicari dulu metode terbaik untuk pembentukan pohon klasifikasi, yakni yang menghasilkan ketepatan klasifikasi pada data *testing* tertinggi. Oleh karena itu, data keseluruhan perlu dibagi menjadi dua bagian yaitu L_1 (data *learning*) dan L_2 (data *testing*). Data *learning* yang masih bersifat heterogen berada dalam suatu ruang atau simpul. Simpul tersebut menjadi simpul utama yang perlu dipilah oleh salah satu variabel prediktor sebagai pemilah utama. Pemilihan pemilah dimulai dengan memeriksa nilai-nilai variabel prediktor pada tiap simpul dengan dua tahap yaitu dengan mencari semua kemungkinan pemilah pada setiap variabel prediktor dan mencari pemilah terbaik dari setiap variabel prediktor.

1. Pemilihan Pemilah

Menurut Breiman dkk (1993), proses pemilahan simpul menjadi dua simpul anak dilakukan dengan mengikuti aturan sebagai berikut.

1. Setiap pemilahan bergantung pada nilai dari satu variabel prediktor saja.
2. Bila X_j berskala kontinu, maka pemilahan yang diperbolehkan adalah $x_j \leq c_i$ dan $x_j > c_i$, dengan $i = 1, 2, \dots, n-1$ dan c_i adalah nilai tengah dari dua nilai amatan sampel berurutan yang berbeda dari variabel X_j . Jika suatu ruang sampel berukuran n dan terdapat sejumlah n nilai amatan sampel yang berbeda pada variabel X_j , maka terdapat $n-1$ kemungkinan pemilahan yang berbeda.
3. Bila X_j merupakan variabel kategorik, maka pemilahan berasal dari semua kemungkinan pemilahan berdasarkan terbentuknya dua simpul yang saling lepas (*disjoint*). Bila kategori berskala nominal bertaraf L , maka akan diperoleh sebanyak $2^{L-1} - 1$ pemilahan yang mungkin. Bila kategori berskala ordinal bertaraf L , maka akan diperoleh sebanyak $L-1$ pemilahan yang mungkin.

Tujuan dilakukan pemilahan adalah untuk mengurangi keheterogenan pada simpul utama dan mendapatkan simpul anak-simpul anak dengan tingkat homogenitas yang tinggi dan dapat mencirikan karakteristik kelas-kelas variabel respon. Untuk mengetahui kondisi tersebut, maka ada implementasi fungsi keheterogenan simpul. Hal ini dapat dilakukan dengan mendefinisikan fungsi keheterogenan simpul (*impurity* atau $i(t)$). Fungsi heterogenitas yang umum digunakan adalah Indeks Gini. Metode ini memiliki kelebihan yaitu proses perhitungan yang sederhana dan relatif cepat, serta mudah dan sesuai untuk diterapkan dalam berbagai kasus (Breiman dkk, 1993). Ide dasar dari Indeks Gini adalah memisahkan kelas dengan anggota paling besar atau kelas terpenting dalam simpul tersebut terlebih dahulu. Pemilah terbaik dipilih dari semua kemungkinan pemilahan pada setiap variabel prediktor berdasarkan pada nilai penurunan

keheterogenan tertinggi. Fungsi Indeks Gini dituliskan dalam persamaan (2.5).

$$i(t) = \sum_{i,j=1} p(i|t)p(j|t), i \neq j \quad (2.5)$$

dengan

$i(t)$ = Fungsi keheterogenan simpul t

$p(i|t)$ = Proporsi kelas i pada simpul t ,

$p(j|t)$ = Proporsi kelas j pada simpul t

Selain Indeks Gini terdapat ukuran lain yang dapat digunakan untuk mendefinisikan fungsi keheterogenan seperti indeks Informasi, indeks Twoing, dan indeks Entropy. Namun pada penelitian ini akan membandingkan antara Indeks Gini dan Indeks Twoing. Berikut merupakan fungsi keheterogenan indeks Twoing.

$$i(t) = \frac{P_L P_R}{4} [\sum |p(j|t_L) - p(j|t_R)|]^2 \quad (2.6)$$

dengan,

$i(t)$ = Fungsi keheterogenan simpul t

P_L = Proporsi pengamatan pada simpul kiri

P_R = Proporsi pengamatan pada simpul kanan

$p(j|t_L)$ = Proporsi pengamatan dari simpul t menuju simpul kiri dengan kelas j

$p(j|t_R)$ = Proporsi pengamatan dari simpul t menuju simpul kanan dengan kelas j

Setelah dilakukan pemilahan dari semua kemungkinan pemilah, maka tahapan berikutnya adalah menentukan pemilihan pemilah terbaik berdasarkan kriteria *goodness of split* ($\phi(s,t)$) untuk mengevaluasi pemilahan oleh suatu pemilah s pada simpul t . *Goodness of split* menunjukkan ukuran penurunan keheterogenan suatu kelas atau kualitas ukuran seberapa baik pemilah s menyaring data menurut kelas agar lebih homogen. *Goodness of split* ($\phi(s,t)$) didefinisikan sebagai penurunan heterogenitas sebagai berikut :

$$\phi((s, t) = \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2.7)$$

dengan,

$\phi((s, t)$ = Nilai *Goodness of split*

$i(t)$ = Fungsi heterogenitas pada simpul t

$p_L i(t_L)$ = Proporsi pengamatan dari simpul t menuju simpul kiri

$p_R i(t_R)$ = Proporsi pengamatan dari simpul t menuju simpul kanan

Pemilah yang menghasilkan $\phi(s, t)$ lebih tinggi merupakan pemilah terbaik karena mampu mereduksi heterogenitas lebih tinggi. Pembentukan struktur pohon dilakukan dengan pencarian semua pemilah yang mungkin pada setiap simpul yang dimulai dari simpul utama t_1 . Suatu pemilah s^* yang mampu menghasilkan nilai penurunan keheterogenan tertinggi diantara semua pemilah s yang mungkin adalah yang terpilih sebagai pemilah simpul utama t_1 menjadi simpul anak t_2 dan t_3 . Konsep tersebut secara sistematis dapat dituliskan dalam persamaan (2.8).

$$\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t_1) \quad (2.8)$$

Kemudian untuk pengembangan pembentukan struktur pohon maka dengan cara dan konsep yang sama dilakukan pemilahan pada simpul t_2 dan t_3 masing-masing secara terpisah, seterusnya sampai ditemukan simpul terminal.

2. Penentuan *Terminal Node*

Suatu simpul t dapat dijadikan *terminal node* atau tidak dapat dipilah lagi tergantung dari kondisi simpul tersebut apabila kondisi simpul tersebut memenuhi salah satu kriteria sebagai berikut.

- a. Hanya ada satu pengamatan ($n=1$) dalam tiap simpul anak atau adanya batasan minimum n pengamatan yang diinginkan peneliti.

- b. Semua pengamatan dalam setiap simpul anak mempunyai distribusi yang identik terhadap variabel prediktor sehingga tidak mungkin untuk dipilih lagi.
- c. Adanya batasan jumlah level atau tingkat kedalaman pohon maksimal yang ditetapkan peneliti.

Apabila struktur pohon telah terbentuk mulai dari simpul utama t_l sampai dengan simpul terminal dimana sudah tidak lagi ditemukan simpul yang perlu dipilah lagi maka itu berarti pohon klasifikasi maksimal telah terbentuk. Pohon klasifikasi maksimal yaitu pohon klasifikasi yang memiliki jumlah simpul paling banyak.

3. Penandaan Label Kelas

Setiap simpul terminal perlu diberi label kelas agar dapat diketahui karakteristik dari klasifikasi pengamatan untuk setiap kelas variabel respon yang terbentuk. Pemberian label kelas pada simpul terminal ditentukan berdasarkan aturan jumlah anggota kelas terbanyak, yaitu jika,

$$p(j_0 | t) = \max_j p(j | t) = \max_j \frac{N_j(t)}{N(t)} \quad (2.9)$$

dengan,

$p(j | t)$ = Proporsi kelas j pada simpul t

$N_j(t)$ = Jumlah pengamatan kelas j pada simpul terminal t

$N(t)$ = Jumlah seluruh pengamatan pada simpul terminal t

Label kelas untuk simpul terminal t adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul t paling kecil sebesar $r(t) = 1 - \max_j p(j | t)$.

2.7.2 Pemangkasan Pohon Klasifikasi

Pohon klasifikasi yang terbentuk dari hasil pemilahan dapat berukuran sangat besar, karena aturan penghentian pohon hanya berdasarkan pada jumlah pengamatan pada simpul akhir yang biasanya ditetapkan sebanyak satu pengamatan saja. Pohon yang besar dapat memunculkan dugaan adanya kasus *overfitting* (nilai

yang dihasilkan melebihi kenyataan yang ada). Sehingga, banyaknya pemilahan yang dilakukan tersebut dapat mengakibatkan makin kecilnya tingkat kesalahan prediksi. Selain itu, bila dalam proses pemilahan diberikan batasan, padahal pemilahan masih layak untuk dilakukan maka akan terjadi kasus *underfitting* (nilai yang dihasilkan di bawah kenyataan). Oleh karena itu, untuk mendapatkan pohon klasifikasi yang layak, perlu dilakukan pemangkasan pohon (*pruning*) yaitu suatu penilaian ukuran pohon tanpa mengorbankan akurasi yang berarti yang dilakukan melalui pengurangan simpul pohon sehingga dicapai ukuran pohon yang layak dan tidak terlalu melebar. Ukuran pohon yang layak dapat dilakukan pemangkasan pohon dengan ukuran *cost complexity* (Breiman dkk, 1993) :

$$R_{\alpha}(T) = R(T) + \alpha |\tilde{T}| \quad (2.10)$$

dengan,

$R_{\alpha}(T)$ = *Cost complexity measure* (ukuran kompleksitas) suatu pohon T pada kompleksitas α

$R(T)$ = *Cost misclassification T tree* (ukuran kesalahan klasifikasi pohon T)

α = *Complexity parameter* (kompleksitas parameter) oleh penambahan satu simpul akhir pada pohon T

$|\tilde{T}|$ = Ukuran banyaknya simpul terminal pada pohon T

Cost complexity pruning menentukan suatu pohon bagian $T(\alpha)$ yang meminimumkan $R_{\alpha}(T)$ pada seluruh pohon bagian atau untuk setiap nilai α . Selanjutnya, dilakukan pencarian pohon bagian $T(\alpha) < T_{\max}$ yang dapat meminimumkan $R_{\alpha}(T)$ yaitu

$$R_{\alpha}(T(\alpha)) = \min_{T < T_{\max}} R_{\alpha}(T)$$

Secara umum tahapan pada proses pemangkasan pohon (*pruning*) adalah sebagai berikut :

1. Membentuk pohon klasifikasi maksimal T_{\max} , kemudian diambil t_R dan t_L yang merupakan simpul anak kanan dan

- simpul anak kiri dari T_{max} yang dihasilkan dari pemilahan simpul induk t .
2. Jika diperoleh dua simpul anak dan simpul induknya yang memenuhi persamaan $R(t) = R(t_L) + R(t_R)$, maka simpul anak t_R dan t_L dipangkas. Hasilnya adalah pohon T_l yang memenuhi kriteria $R(T_l) = R(T_{max})$.
 3. Ulangi langkah di atas sampai tidak ada lagi pemangkasan yang mungkin. Hasil proses pemangkasan di atas adalah suatu barisan menurun dan tersarang dari pohon bagian yaitu $T_l > T_2 > T_3 > \dots > \{t_l\}$ dengan $T_l < T_{max}$ dan suatu barisan menaik dari parameter *cost complexity*, yaitu $\alpha_1 = 0 < \alpha_2 < \alpha_3 < \dots$ sebagaimana untuk $\alpha_k \leq \alpha < \alpha_{k+1}$, T_k adalah sub pohon terkecil dari T_{max} yang meminimumkan $R_\alpha(T)$.

2.7.3 Penentuan Pohon Klasifikasi Optimal

Ukuran pohon yang terlalu besar akan menyebabkan nilai *cost complexity* yang tinggi karena struktur data yang digambarkan cenderung kompleks sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti yang cukup kecil. Bila $R(T)$ dipilih sebagai penduga terbaik, maka akan cenderung dipilih pohon yang besar, sebab pohon yang semakin besar akan membuat nilai $R(T)$ semakin kecil. Terdapat dua macam penduga untuk mendapatkan pohon klasifikasi optimal yaitu penduga sampel uji (*test sample estimate*) dan penduga validasi silang lipat v (*cross validation v-fold estimate*).

a. Penduga Sampel Uji (*Test Sample Estimation*)

Penduga sampel uji dilakukan jika data berukuran besar. Prosedur *test sample estimation* diawali dengan membagi data penelitian menjadi dua bagian yaitu L_1 (*data learning*) dan L_2 (*data testing*). Pengamatan dalam L_1 digunakan untuk membentuk pohon T , sedangkan pengamatan dalam L_2 digunakan untuk

menduga $R^s(T_t)$. Persamaan *test sample estimation* adalah sebagai berikut.

$$R^{ts}(T_k) = \frac{1}{N^{(2)}} \sum_{i,j} C(i|j) N_{ij}^{(2)} \quad (2.11)$$

dengan

$R^{ts}(T_k)$ = Total proporsi kesalahan *test sample estimation*

$N^{(2)}$ = Jumlah pengamatan dari L_2

$C(i|j)$ = Jumlah proporsi ke- i dan ke- j dari L_2

$N_{ij}^{(2)}$ = Jumlah kelas ke- j dalam L_2 dimana klasifikasi prediksinya adalah kelas ke- i

Pohon klasifikasi optimal yang dipilih yaitu T_k dengan

$$R^{ts}(T_k) = \min_t R^{ts}(T_k) .$$

b. Penduga Validasi Silang *V-Fold (Cross Validation V-Fold Estimation)*

Penduga pengganti ini sering dilakukan apabila pengamatan yang ada tidak cukup besar. *Cross validation* membagi data secara acak menjadi V subset yang berukuran relatif sama. salah satu subset dicadangkan sebagai data *testing* dan subset-subset sisanya digabung dijadikan sebagai data *learning* dalam prosedur pembentukan model. Seluruh prosedur pembentukan model diulang V kali, dengan subset berbeda dari data setiap kali melakukan pembentukan pohon (Lewis dan Roger, 2000). Penduga ini bisa dicoba untuk menghindari *overlapping* pada data *testing*. Nilai V yang sering dipakai dan dijadikan standar adalah 10. Karena hasil dari berbagai percobaan ekstensif dan pembuktian teoritis, menunjukkan bahwa *cross validation 10-fold* adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat.

Cross validation v-fold estimation untuk T_k yang menggunakan pengamatan L dalam membentuk deretan pohon $\{T_k\}$ adalah sebagai berikut.

$$R^{cv}(T_k(\alpha)) = \frac{1}{N} \sum_{i,j} C(i|j) N_{ij} \quad (2.12)$$

dengan,

$R^{cv}(T_k(\alpha))$ = Total proporsi *T Cross validation v-fold estimation*

$C(i|j)$ = Jumlah proporsi ke-*i* dan ke-*j* dari keseluruhan data pengamatan

N_{ij} = Jumlah kelas ke-*i* dan ke-*j* dari keseluruhan data pengamatan

Pohon klasifikasi optimal yang dipilih yaitu T_k dengan

$$R^{cv}(T_k) = \min_k R^{cv}(T_k)$$

2.7.4 Ukuran Ketepatan Klasifikasi

Beberapa cara yang dapat digunakan untuk mengukur ketepatan klasifikasi diantaranya adalah melalui perhitungan *sensitivity*, *Apparent Error Rate* (APER) dan *total accuracy rate* (1-APER). *Sensitivity* adalah ukuran ketepatan dari kejadian yang diinginkan. *Specificity* adalah ukuran yang menyatakan persentase dari kejadian yang tidak diinginkan. Menurut Johnson dan Wichern (2007), *Apparent Error Rate* (APER) merupakan proporsi observasi yang diprediksi secara tidak benar (ukuran kesalahan klasifikasi total). *Total accuracy rate* atau (1-APER) merupakan proporsi observasi yang diprediksi secara benar (ukuran ketepatan klasifikasi total). *Crosstab* untuk menghitung ketepatan klasifikasi ditunjukkan dalam Tabel 2.6.

Tabel 2. 6 *Crosstab* Ketepatan Klasifikasi

Kelas Observasi Y	Kelas Prediksi Y		Total
	1	2	
1	n_{11}	n_{12}	$N_{1.}$
2	n_{21}	n_{22}	$N_{2.}$
Total	$N_{.1}$	$N_{.2}$	N

dengan,

n_{11} = Frekuensi variabel Y pada kategori 1 yang tepat diprediksikan sebagai variabel Y kategori 1

n_{21} = Frekuensi variabel Y pada kategori 2 yang salah diprediksikan sebagai variabel Y kategori 1

n_{12} = Frekuensi variabel Y pada kategori 1 yang salah diprediksikan sebagai variabel Y kategori 2

n_{22} = Frekuensi variabel Y pada kategori 2 yang tepat diprediksikan sebagai variabel Y kategori 2

N_1 = Frekuensi variabel Y pada kategori 1

N_2 = Frekuensi variabel Y pada kategori 2

$N_{.1}$ = Jumlah prediksi dari variabel Y pada kategori 1

$N_{.2}$ = Jumlah prediksi dari variabel Y pada kategori 2

N = Frekuensi total/prediksi.

Berikut adalah perhitungan untuk *sensitivity*, *specificity*, APER, dan *total accuracy rate*.

$$Sensitivity = \frac{n_{11}}{N_{.1}} \quad (2.13)$$

$$Specificity = \frac{n_{22}}{N_{.2}} \quad (2.14)$$

$$Apparent \ Error \ Rate \ (APER) = \frac{n_{21} + n_{12}}{N} \quad (2.15)$$

$$Total \ accuracy \ rate = 1 - APER = \frac{n_{11} + n_{22}}{N} \quad (2.16)$$

2.8 Penyebaran Penyakit Malaria di Indonesia

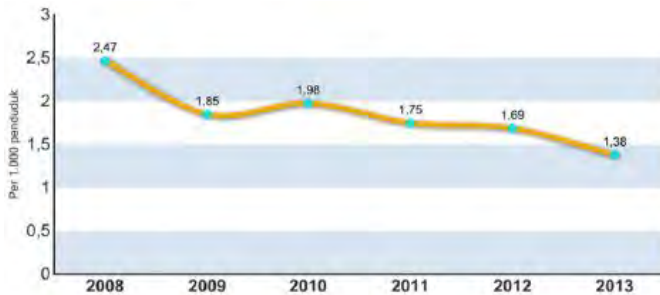
Malaria merupakan penyakit menular yang menjadi perhatian global. Penyakit ini masih merupakan masalah kesehatan masyarakat karena sering menimbulkan KLB, berdampak luas terhadap kualitas hidup dan ekonomi, serta dapat

mengakibatkan kematian. Penyakit ini dapat bersifat akut, laten atau kronis (Kementerian Kesehatan RI, 2013)

Malaria disebabkan parasit *Plasmodium sp.* dan dapat bersifat kronis karena parasit dapat bersembunyi dalam tubuh pejamu dan menimbulkan manifestasi klinis sewaktu-waktu, ketika daya tahan tubuh pejamu menurun. Spesies parasit yang dominan menyebabkan malaria di Indonesia sampai saat ini masih *Plasmodium falciparum* dan *Plasmodium vivax*, yang juga diketahui dapat menimbulkan malaria berat. Malaria berat tersebut terutama disebabkan infeksi *Plasmodium falciparum* karena menyerang otak dan komplikasi lainnya, seperti kegagalan organ multipel pada kasus yang sangat parah (Kementerian Kesehatan RI, 2013).

Malaria merupakan salah satu masalah kesehatan masyarakat yang dapat menyebabkan kematian terutama pada kelompok risiko tinggi yaitu bayi, anak balita, ibu hamil, selain itu malaria secara langsung menyebabkan anemia dan dapat menurunkan produktivitas kerja. Penyakit ini juga masih endemis di sebagian besar wilayah Indonesia (Kementerian Kesehatan RI, 2011)

Banyak hal yang sudah maupun sedang dilakukan baik dalam skala global maupun nasional dalam rangka pengendalian penyakit malaria. Malaria merupakan salah satu indikator dari target Pembangunan Milenium (MDGs), dimana ditargetkan untuk menghentikan penyebaran dan mengurangi kejadian insiden malaria pada tahun 2015 yang dilihat dari indikator menurunnya angka kesakitan dan angka kematian akibat malaria. Upaya penanggulangan penyakit malaria di Indonesia sejak tahun 2007 dapat dipantau dengan menggunakan indikator *Annual Parasite Incidence* (API) (Kementerian Kesehatan RI, 2014). Gambar 2.3 menunjukkan API dari tahun 2008 hingga 2013 mengalami penurunan dari 2,47 per 1.000 penduduk menjadi 1,38 per 1.000 penduduk.



Gambar 2. 4 Annual Parasite Incidence (API) per 1,000 Penduduk di Indonesia Tahun 2008-2013

2.9 Faktor-Faktor yang Mempengaruhi Penyebaran Penyakit Malaria di Indonesia

Penyebaran penyakit malaria disebabkan oleh beberapa faktor antara lain adalah sebagai berikut.

1. Wilayah Tempat Tinggal

Menurut Harijanto (2000), dalam masyarakat modern sering dibedakan antara masyarakat perdesaan (*rural community*) dan masyarakat perkotaan (*urban community*). Secara umum pada daerah perkotaan kasus malaria cenderung menurun, hal ini disebabkan oleh membaiknya sanitasi lingkungan dan meningkatnya pengetahuan serta perilaku sehat masyarakat perkotaan. Sedangkan daerah pinggiran atau perdesaan, lingkungan fisiknya lebih memungkinkan perkembangan vektor nyamuk, misalnya dengan adanya rawa, genangan air di hutan, sawah, dan tambak. Prevalensi malaria berdasarkan karakteristik tempat tinggal di perdesaan (8%) dua kali prevalensi di perkotaan (4%) (Kementrian Kesehatan RI, 2011).

2. Jenis Kelamin Kepala Keluarga

Insiden berbagai penyakit di antara jenis kelamin berbeda terutama disebabkan paparan terhadap agen bagi setiap jenis kelamin berbeda. Jenis pekerjaan pria dan wanita berbeda, pembagian pekerjaan secara sosial antara wanita dan laki-

laki menyebabkan perbedaan paparan yang diterima orang, sehingga penyakit yang dialami berbeda pula. Beberapa penelitian menunjukkan bahwa perempuan mempunyai respon imun yang lebih kuat dibandingkan dengan laki-laki (Harijanto, 2000).

3. Umur

Banyak diketahui bahwa ada penyakit yang disebut penyakit anak, penyakit orang tua, dan sebagainya. Umur dapat menentukan seseorang rentan terkena malaria atau tidak karena terkait dengan kekebalan, hormonal dan ketahanan tubuh. Umur merupakan faktor yang berpengaruh signifikan terhadap kejadian sakit malaria. (Saikhu, Budianto, dan Yuliani, 2009).

4. Pendidikan terakhir Kepala Keluarga

Tingkat pendidikan terakhir seseorang dapat berpengaruh terhadap penyakit malaria. Semakin tinggi tingkat pendidikan seseorang maka pengetahuan yang dimiliki semakin banyak, sehingga berbagai bentuk pencegahan terhadap faktor penyebab malaria dapat dilakukan. Orang tua yang memiliki tingkat pendidikan tinggi berpengaruh terhadap hubungan kondisi fisik bangunan rumah dengan kejadian malaria pada anak sebagai faktor *confounding* yang bersifat protektif mengurangi resiko untuk terkena malaria sebesar 0,47 kali lebih kecil dibandingkan orang tua yang memiliki tingkat pendidikan rendah (Frits, 2003).

5. Pekerjaan Kepala Keluarga

Jenis pekerjaan mempunyai hubungan dengan malaria. Pekerjaan tertentu merupakan faktor risiko untuk terkena malaria, misalnya tukang kebun, petani, nelayan, buruh, dan lain-lain. Jenis pekerjaan tersebut memberi peluang untuk kontak dengan nyamuk (Achmadi, 2008; Soemirat, 2000). Anggota rumah tangga yang tidak bekerja mempunyai peluang yang lebih besar 1,256 kali didiagnosa malaria

dibandingkan dengan anggota rumah tangga yang bekerja sebagai PNS (Ekayani, 2011).

6. Status Ekonomi

Status ekonomi suatu rumah tangga terdiri dari 5 macam, yaitu terbawah, menengah bawah, menengah, menengah atas, dan teratas. Status ekonomi suatu rumah tangga dapat berhubungan dengan kejadian malaria. Karena secara umum, seseorang dengan status ekonomi yang lebih tinggi akan memiliki kehidupan yang layak dan cenderung untuk tidak terkena malaria. Hal tersebut dapat dibuktikan pada laporan Riskesdas 2013, insiden terdiagnosa malaria pada kuintil yang lebih rendah (terbawah) terdiagnosa malaria sebanyak 0,8% sedangkan kuintil menengah bawah sebesar 0,4%, serta pada kuintil menengah hingga kuintil teratas terdapat insiden malaria sebanyak 0,2%.

7. Jumlah Pemakaian Air

Menurut laporan Riskesdas 2013, pemakaian air per orang per hari oleh rumah tangga di Indonesia, pada umumnya berjumlah antara 50 sampai 99,9 liter (28,3%), dan antara 100 sampai 300 liter (40%). Menurut karakteristik, proporsi rumah tangga dengan pemakaian air kurang dari 20 liter per orang per hari di perdesaan lebih tinggi (5,8%) dibandingkan di perkotaan (4,0%), sebaliknya proporsi rumah tangga jumlah pemakaian air per orang per hari 20 liter atau lebih di perkotaan lebih tinggi (95,9%) dibandingkan dengan di perdesaan (94,2%). Berdasarkan provinsi, proporsi rumah tangga dengan jumlah pemakaian air per orang per hari kurang dari 20 liter tertinggi adalah Nusa Tenggara Timur (30,4%) diikuti Papua (22,5%). Papua dan NTT termasuk dalam lima provinsi dengan insiden dan prevalensi malaria tertinggi di Indonesia, jumlah pemakaian air diduga menjadi salah satu faktor yang mempengaruhi penyakit malaria.

8. Pengolahan Air Sebelum Diminum

Proporsi rumah tangga yang mengolah air sebelum di minum di Indonesia sebesar 70,1 persen. Rumah tangga yang melakukan pengolahan air sebelum diminum dengan cara dimasak sebesar 96,5%. Cara pengolahan lainnya adalah dengan dijemur di bawah sinar matahari/*solar disinfection* (2,3%), menambahkan larutan tawas (0,2%), disaring dan ditambah larutan tawas (0,2%) dan disaring saja (0,8%). Menurut karakteristik, proporsi rumah tangga yang melakukan pengolahan air sebelum diminum dengan cara pemanasan/dimasak, di perkotaan (96,5%) hampir sama dengan di perdesaan (96,6%) (Kementerian Kesehatan RI, 2013).

9. Pelayanan Kesehatan Gratis

Berdasarkan Laporan Riskesdas 2013, pemerintah memberikan pelayanan kesehatan gratis bagi masyarakat miskin di seluruh Indonesia melalui berbagai program diantaranya Jaminan Kesehatan Masyarakat (Jamkesmas), Jaminan Kesehatan Daerah (Jamkesda), Program Keluarga Harapan (PKH) dan Kartu Sehat. Selain itu, masyarakat miskin yang tidak tercatat dalam database Jamkesmas, Jamkesda, PKH dan Kartu Sehat juga mendapatkan pelayanan kesehatan gratis dengan menggunakan surat keterangan tidak mampu (SKTM).

10. Sumber Air Utama

Jenis sumber air utama terbagi menjadi 2, yaitu sumber air terlindung dan tidak terlindung. Untuk akses terhadap sumber air minum digunakan kriteria JMP WHO - Unicef tahun 2006. Menurut kriteria tersebut, rumah tangga memiliki akses ke sumber air minum *improved* adalah rumah tangga dengan sumber air minum dari air ledeng/PDAM, sumur bor/pompa, sumur gali terlindung, mata air terlindung, penampungan air hujan, dan air kemasan (hanya jika sumber air untuk keperluan rumah tangga lainnya *improved*).

Proporsi rumah tangga yang memiliki akses terhadap sumber air minum *improved* di Indonesia adalah sebesar 66,8 persen. Lima provinsi dengan akses sumber air terlindung terendah adalah Kepulauan Riau (24,0%), Kalimantan Timur (35,2%), Bangka Belitung (44,3%), Riau (45,5%), dan Papua (45,7%). (Kementerian Kesehatan RI, 2013). Papua merupakan provinsi dengan insiden dan prevalensi malaria tertinggi di Indonesia, sehingga jenis sumber air utama diduga menjadi salah satu faktor yang mempengaruhi penyakit malaria.

11. Pencegahan Gigitan Nyamuk

Terdapat berbagai cara pencegahan nyamuk, diantaranya adalah menggunakan obat nyamuk, pemakaian kelambu dan pemakaian kawat kasa. Anggota rumah tangga yang menggunakan kelambu memiliki peluang 0,794 kali lebih kecil untuk terkena malaria dibandingkan dengan yang tidak menggunakan kelambu (Ekayani, 2011).

12. Lingkungan Kumuh

Lingkungan tempat tinggal yang kumuh serta sanitasi buruk biasanya menjadi penyebab utama berjangkitnya penyakit malaria karena parit-parit, persawahan, empang, dan genangan air merupakan tempat bersarangnya nyamuk *Anopheles* (Soemirat, 2000).

13. Kepadatan Hunian

Kepadatan hunian merupakan salah satu persyaratan rumah sehat. Dalam Keputusan Menteri Kesehatan no 829/Menkes/SK/VII/1999 tentang Persyaratan Kesehatan Perumahan, disebutkan bahwa kepadatan hunian kurang dari 8orang/m² dikategorikan sebagai tidak padat. Proporsi rumah tangga di Indonesia yang termasuk ke dalam kriteria tidak padat sebesar 86,6%. Lima provinsi terendah tidak padat (<8orang/m²) adalah Papua (55,0%), NTT (64,0%), DKI Jakarta (68,3%), Gorontalo (69,0%), dan Maluku (72,7%). Papua, NTT, dan Maluku termasuk dalam lima provinsi dengan insiden dan prevalensi malaria tertinggi di Indonesia

(Kementerian Kesehatan RI, 2013). Sehingga faktor kepadatan hunian diduga berpengaruh terhadap kejadian penyakit malaria.

2.9 Penelitian Sebelumnya

Penelitian sebelumnya mengenai penyakit malaria telah dilakukan oleh beberapa orang, antara lain oleh Ekayani (2011) yang melakukan pemodelan antara penyakit malaria dengan faktor-faktor yang diduga dapat mempengaruhi penyakit malaria di Provinsi Papua Barat dengan menggunakan regresi logistik biner menghasilkan kesimpulan bahwa faktor-faktor yang berpengaruh signifikan adalah faktor pekerjaan, saluran pembuangan limbah, tempat penampungan limbah, penggunaan kelambu, sarana penampungan air minum, pemanfaatan posyandu, dan pemanfaatan POD/WOD.

Penelitian lain oleh Susilowati (2014) yang meneliti tentang prevalensi malaria di Provinsi Maluku Utara, Maluku, Papua Barat dan Papua menggunakan regresi *ordinary least square* (OLS) dan *robust* menghasilkan kesimpulan faktor-faktor yang berpengaruh yaitu persentase rumah tangga tidak memiliki saluran pembuangan air limbah dan persentase rumah tangga mengakses air bersih.

Penelitian yang dilakukan oleh Lestari (2014) tentang faktor-faktor yang mempengaruhi penyakit malaria pada ibu hamil di Nusa Tenggara Barat, Nusa Tenggara Timur, Maluku, Maluku Utara, Papua dan Papua Barat menggunakan metode regresi logistik biner. Penelitian ini menghasilkan faktor-faktor yang memiliki pengaruh signifikan terhadap penyakit malaria pada ibu hamil adalah rumah panggung, atap ijuk/rumbia, atap seng, dan lantai semen plesteran retak.

Salah satu penelitian yang menggunakan metode CART dilakukan oleh Irawan (2014) mengenai klasifikasi status HIV/AIDS di LSM Orbit Surabaya menghasilkan kesimpulan bahwa ketepatan klasifikasi status HIV/AIDS sebesar 65% dan hasil validasi pada data *testing* sebesar 73,68%. Penelitian lain

dilakukan oleh Seftiana (2014) tentang klasifikasi rumah tangga sangat miskin di kabupaten Jombang dengan metode RF-CART menghasilkan tingkat akurasi sebesar 65,5% untuk data *learning* dan 62,8% untuk data *testing*.

Penelitian dengan menggunakan metode SMOTE pernah dilakukan oleh Trapsilasiwi (2014) mengenai kasus kanker payudara dan kanker serviks dan menghasilkan kesimpulan bahwa pada klasifikasi kanker payudara menghasilkan tingkat akurasi sebesar 89,2% sebelum menggunakan SMOTE dan 96,8% sesudah menggunakan SMOTE. Sedangkan pada klasifikasi kanker serviks, menghasilkan ketepatan akurasi sebesar 40,4% sebelum menggunakan SMOTE dan 59,3% sesudah menggunakan SMOTE.

BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

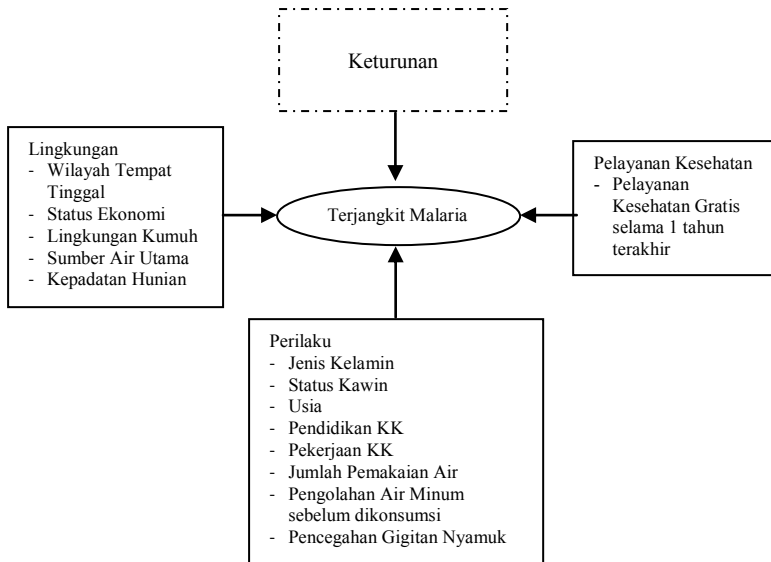
Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari hasil Riset Kesehatan Dasar (Riskesdas) tahun 2013. Upaya penjaminan mutu data Riskesdas 2013 adalah melakukan uji coba instrumen dan validasi. Uji coba bertujuan untuk menilai keabsahan instrumen antara lain mendapatkan kuesioner yang sesuai dengan tujuan dalam Riskesdas, menentukan kelayakan dari peralatan yang akan digunakan serta manajemen pengumpulan data. Uji coba dilakukan oleh peneliti Badan Litbangkes, akademisi, dan organisasi profesi. Sedangkan validasi Riskesdas 2013 adalah kunjungan ulang sub sampel Riskesdas 2013 yang dilakukan sebagai salah satu bagian dari *quality assurance* untuk menjamin kualitas data Riskesdas 2013. Validasi dilakukan oleh tim universitas (Kementerian Kesehatan RI, 2013). Data yang digunakan dalam penelitian ini berupa data mentah tentang penyakit malaria di Provinsi Papua Barat dengan unit penelitian adalah rumah tangga yang jumlahnya mencapai 3.836 rumah tangga. Adapun surat keterangan data yang diperoleh dari Riskesdas terlampir di Lampiran F. Struktur data dalam penelitian ini adalah sebagai berikut.

Tabel 3. 1 Struktur Data Penelitian

No.	Y	X_1	X_2	\dots	X_{14}
1	y_1	x_{11}	x_{21}	\dots	$x_{14,1}$
2	y_2	x_{12}	x_{22}	\dots	$x_{14,2}$
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
N	y_n	x_{1n}	x_{2n}	\dots	$x_{14,n}$

3.2 Kerangka Konsep Penelitian

Kerangka konsep dalam penelitian ini mengacu pada kerangka konsep status kesehatan Blum (1976). Berikut adalah kerangka konsep yang digunakan dalam penelitian ini.



Gambar 3. 1 Kerangka Konsep Penelitian (Pengembangan Faktor Status Kesehatan Blum)

(Sumber : Notoadmodjo, 2010)

3.3 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari variabel respon (Y) dan prediktor (X) sesuai dengan survei Riset Kesehatan Dasar 2013. Berikut adalah identifikasi untuk variabel respon dan variabel prediktor.

A. Variabel Respon

Variabel respon yang digunakan dalam penelitian ini adalah status rumah tangga terhadap penyakit malaria.

Tabel 3. 2 Variabel Respon

Variabel	Kategori	Skala
Status rumah tangga terhadap penyakit malaria (Y)	1 = rumah tangga dengan anggota rumah tangga terjangkit malaria	Nominal
	2 = rumah tangga dengan anggota rumah tangga tidak terjangkit malaria	

Rumah tangga dikatakan terjangkit malaria jika didalam rumah tangga tersebut terdapat minimal satu anggota rumah tangga (ART) pernah dinyatakan positif terjangkit malaria secara medis dalam satu bulan terakhir.

B. Variabel Prediktor

Variabel prediktor yang digunakan dalam penelitian ini dibagi kedalam dua kategori, yakni karakteristik demografi dan sosial ekonomi serta faktor risiko. Berikut adalah identifikasi dari setiap variabel dalam penelitian ini.

- a. Karakteristik Demografi dan Sosial Ekonomi Responden (Kepala Rumah Tangga)

Tabel 3.3 Variabel Penelitian Demografi dan Sosial Ekonomi Responden

	Variabel	Definisi Operasional	Kategori	Skala
X ₁	Wilayah Tempat Tinggal	Klasifikasi dibedakan antara masyarakat perdesaan (<i>rural community</i>) dan masyarakat perkotaan (<i>urban community</i>)	1=Perkotaan 2=Perdesaan	Nominal
X ₂	Jenis Kelamin	Tanda seks sekunder pada diri seseorang	1=Laki-laki 2=Perempuan	Nominal
X ₃	Status Kawin	Status kawin responden saat diwawancarai	1 = Belum Menikah 2 = Menikah 3 = Hidup Bersama 4 = Cerai Hidup 5 = Hidup Terpisah 6 = Cerai mati	Nominal
X ₄	Umur	Umur obyek yang dihitung dalam tahun, yang dihitung berdasarkan waktu ulang tahun terakhir	-	Rasio

Tabel 3. 3 Variabel Penelitian Demografi dan Sosial Ekonomi Responden (Lanjutan)

No	Variabel	Definisi Operasional	Kategori	Skala
X ₅	Pendidikan Terakhir	Tingkat pendidikan tertinggi yang sudah ditamatkan oleh responden	1=Tidak/belum Pernah Sekolah	Ordinal
			2 = Tidak Tamat SD/MI 3 = Tamat SD/MI 4 = Tamat SLTP/MTS 5 = Tamat SLTA/MA 6 = Tamat D1/D2/D3 7 = Tamat PT	
X ₆	Pekerjaan	Status pekerjaan responden saat diwawancarai	1=PNS/TNI/POLRI/BUMD 2=Pegawai Swasta 3=Wiraswasta 4=Petani 5=Nelayan 6=Buruh 7=Lainnya 8=Tidak Bekerja	Nominal
X ₇	Status Ekonomi	Status ekonomi suatu rumah tangga terdiri dari 5 macam, yaitu terbawah, menengah bawah, menengah, menengah atas, dan teratas.	1 = Kuintil 1-3 2 = Kuintil 4-5	Ordinal

b. Faktor-Faktor Risiko Terjangkit Penyakit Malaria

Tabel 3. 4 Variabel Penelitian Faktor Resiko

No	Variabel	Definisi Operasional	Kategori	Skala
X ₈	Jumlah Pemakaian Air untuk Keperluan Rumah Tangga	Jumlah pemakaian air untuk seluruh keperluan rumah tangga dalam sehari (liter)	-	Rasio

Tabel 3. 4 Variabel Penelitian Faktor Resiko (Lanjutan)

No	Variabel	Definisi Operasional	Kategori	Skala
X ₉	Melakukan Pengolahan Air Minum Sebelum dikonsumsi	Adanya pengolahan air minum sebelum diminum	1=Ya 2=Tidak	Nominal
X ₁₀	Pelayanan Kesehatan Gratis	Adanya pelayanan kesehatan gratis dalam 1 tahun terakhir	1 = Ya 2 = Tidak	Nominal
X ₁₁	Jenis Sumber Air Utama	Jenis sumber air utama untuk kebutuhan minum	1 =Air tidak terlindungi 2 = Air terlindungi	Nominal
X ₁₂	Adanya Pencegahan Gigitan Nyamuk	Adanya salah satu upaya pencegahan terhadap gigitan nyamuk	1 = Tidak Ada 2 = Ada	Nominal
X ₁₃	Tinggal di Daerah Kumuh	Kondisi lingkungan tempat responden tinggal dengan criteria mempunyai saluran limbah (got) dan mengalir, mempunyai tempat penampungan sampah dan sampah tidak berserakan	1=Ya 2=Tidak	Nominal
X ₁₄	Kepadatan Hunian	Kepadatan hunian dalam satu bangunan rumah	1=< 8m ² /orang (padat) 2=> 8m ² /orang (tidak padat)	Nominal

3.5 Langkah Analisis Data

Langkah-langkah analisis data yang dilakukan pada penelitian ini adalah sebagai berikut.

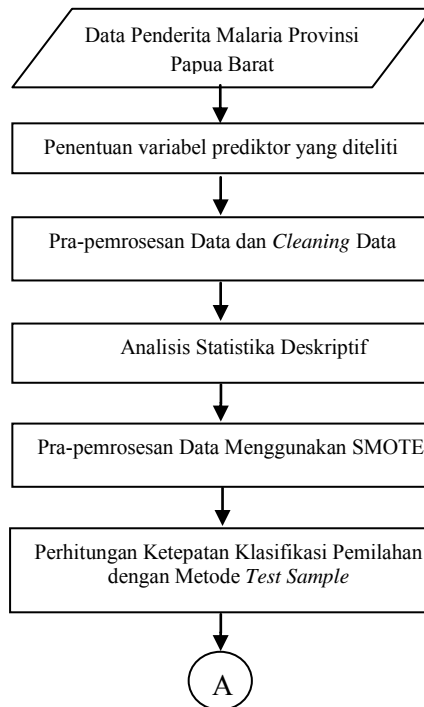
1. Mendapatkan data jumlah kasus malaria di Provinsi Papua Barat dari hasil Riset Kesehatan Dasar 2013
2. Memilih variabel yang akan digunakan dalam penelitian, sebagaimana diuraikan pada subbab 3.3.
3. Pra-pemrosesan data yang sudah terkumpul (3.836 data) dengan melakukan pengkodean data pada setiap variabel

bertipe kategorik sesuai dengan pengkategorian yang telah ditetapkan. Selain itu juga melakukan *cleaning* data terhadap data-data pengamatan yang banyak ditemukan kosong atau tidak diisi secara lengkap oleh responden sehingga menyebabkan banyak informasi yang kurang dari unit pengamatan tersebut.

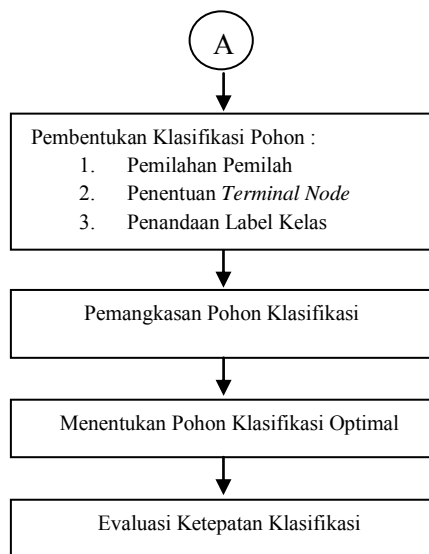
4. Mendeskripsikan gambaran data rumah tangga dengan malaria di Provinsi Papua Barat berdasarkan variabel-variabel yang ada.
5. Menyeimbangkan data menggunakan algoritma *synthetic minority oversampling technique* (SMOTE). Algoritma SMOTE yang digunakan adalah sebagai berikut.
 - a. Menentukan kelas data minor
 - b. Menentukan data yang akan direplikasi pada kelas minor (x_i).
 - c. Menentukan data dengan jarak terdekat dari data yang akan direplikasi dalam satu kelas yang sama (x_{km}).
 - d. Menghitung data sintetis.
6. Membagi data dari hasil langkah 5 menjadi dua bagian yaitu data *learning* dan *testing*. Data dibagi menurut kombinasi data *learning* dan data *testing* dengan proporsi sebesar 75%:25%, 80%:20%, 85%:15%, 90%:10%, 95%:5%. Masing-masing kombinasi proporsi tersebut diolah untuk mencobakan alternatif metode pemilahan *test sample estimation*. Sehingga diperoleh suatu nilai ketepatan klasifikasi dan banyak simpul terminal yang terbentuk dari masing-masing kombinasi proporsi data *learning* dan data *testing* tersebut.
7. Melakukan analisis klasifikasi dengan metode *Classification and Regression Tree* (CART).
 - a. Melakukan pembentukan pohon klasifikasi yaitu dengan tahap pemilihan pemilah menggunakan indeks Gini dan indeks Twoing, penentuan *terminal node*, dan penandaan label kelas.

- b. Melakukan pemangkasan pohon klasifikasi yang maksimal sampai diperoleh ukuran pohon klasifikasi yang paling kecil dengan ukuran *cost complexity minimum*.
- c. Menentukan pohon klasifikasi optimal menggunakan penduga sampel uji (*test sample estimation*)
- d. Mengevaluasi ketepatan klasifikasi untuk mendapatkan nilai *sensitivity*, APER, dan *total accuracy rate* yang dihasilkan oleh data *learning* dan data *testing* dari pohon klasifikasi optimal.

Tahapan metode analisis data secara singkat diuraikan oleh diagram alir pada Gambar 3.2 berikut.



Gambar 3. 2 Diagram Alir Analisis Data



Gambar 3. 2 Diagram Alir Analisis Data (Lanjutan)

BAB IV

ANALISIS DAN PEMBAHASAN

Pada bab ini dilakukan ulasan tentang hasil pengolahan data dan analisis data untuk menjawab permasalahan penelitian. Pada subbab 4.1 diberikan penjelasan statistika deskriptif tentang karakteristik rumah tangga di Provinsi Papua Barat tahun 2013 guna memberikan gambaran mengenai unit analisis yang diteliti. Pada subbab 4.2 diberikan hasil uji independensi guna mengetahui hubungan antara status rumah tangga terhadap penyakit malaria dengan faktor-faktor yang mempengaruhinya. Kemudian pada subbab 4.3 diberikan penjelasan analisis klasifikasi rumah tangga terhadap penyakit malaria di Provinsi Papua Barat dengan pendekatan pohon klasifikasi CART.

4.1 Karakteristik Rumah Tangga di Provinsi Papua Barat pada Tahun 2013

Pada penelitian ini, data rumah tangga di Provinsi Papua Barat pada tahun 2013 disajikan dalam bentuk *pie chart*, diagram batang dan tabel kontingensi untuk menggambarkan karakteristik rumah tangga di Provinsi Papua Barat tahun 2013 berdasarkan wilayah tempat tinggal, jenis kelamin kepala keluarga, status kawin kepala keluarga, umur kepala keluarga, pendidikan terakhir kepala keluarga, pekerjaan kepala keluarga, status ekonomi, jumlah pemakaian air untuk keperluan rumah tangga dalam sehari, pengolahan air minum, pelayanan kesehatan gratis, jenis sumber air utama, adanya pencegahan gigitan nyamuk, daerah kumuh, dan kepadatan hunian.

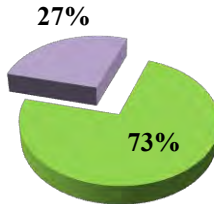
4.1.1 Status Rumah Tangga terhadap Penyakit Malaria

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan survei Riskesdas 2013, terdapat sebanyak 3.836 rumah tangga yang tersebar di 11 kabupaten/kota. Gambar 4.1 menunjukkan banyaknya rumah tangga yang terjangkit penyakit malaria dan tidak terjangkit malaria.

Gambar 4.1 memberikan informasi bahwa sebanyak 1019 rumah tangga (27 persen) dari jumlah keseluruhan rumah tangga di Provinsi Papua Barat terjangkit penyakit malaria, sedangkan

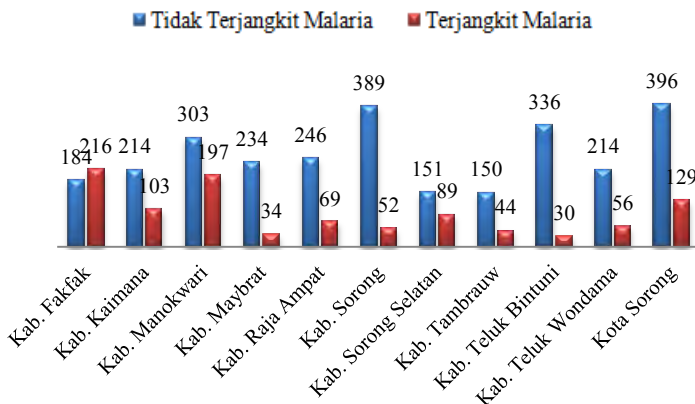
sisanya yakni sebanyak 2817 rumah tangga (73 persen) tidak terjangkit penyakit malaria.

■ Tidak Terjangkit Malaria ■ Terjangkit Malaria



Gambar 4. 1 Karakteristik Rumah Tangga Penderita Penyakit Malaria

Karakteristik rumah tangga di setiap kabupaten/kota di Provinsi Papua Barat pada tahun 2013, ditunjukkan oleh Gambar 4.2.



Gambar 4. 2 Karakteristik Rumah Tangga per Kelas Pada Tiap Kabupaten/Kota di Provinsi Papua Barat

Berdasarkan Gambar 4.2, diperoleh informasi bahwa kabupaten dengan rumah tangga terjangkit penyakit malaria tertinggi adalah Kabupaten Fakfak, yaitu sebanyak 216 rumah tangga dari total rumah tangga sebanyak 400 rumah tangga. Sedangkan kabupaten dengan rumah tangga terjangkit penyakit

malaria terendah adalah Kabupaten Teluk Bintuni, yaitu sebanyak 30 rumah tangga dari total rumah tangga sebanyak 366 rumah tangga.

Pada kenyataannya, data yang diperoleh dari survei Risesdas 2013 tentang rumah tangga di Provinsi Papua Barat tidak semua unit analisis yang diteliti mampu memberikan informasi secara jelas, tepat, dan lengkap sebagaimana yang diharapkan. Karena dari keempat belas variabel prediktor yang diteliti banyak ditemukan data-data *missing*. Oleh karena itu, sebelum dilakukan analisis klasifikasi terlebih dahulu dilakukan pra-pemrosesan data. Unit analisis yang mengandung informasi tidak lengkap tidak diikuti sertakan dalam analisis berikutnya. Sehingga diperoleh data pengamatan yang siap dianalisis lebih lanjut ada sebanyak 3.373 unit rumah tangga. Data yang sudah siap tersebut kemudian di analisis statistik deskriptif dengan cara menghitung rata-rata, standar deviasi, nilai minimum dan maksimum untuk variabel prediktor berskala rasio dan tabel kontingensi untuk variabel yang berskala nominal. Berikut adalah hasil statistika deskriptifnya.

4.1.2 Wilayah Tempat Tinggal

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan wilayah tempat tinggal rumah tangga dapat dilihat pada Tabel 4.1.

Berdasarkan Tabel 4.1, dapat diketahui bahwa mayoritas rumah tangga di Provinsi Papua Barat yang tinggal di daerah perdesaan yaitu sebanyak 2381 rumah tangga, dengan 575 rumah tangga diantaranya (sebesar 17 persen dari total rumah tangga keseluruhan) terjangkit penyakit malaria dan sisanya yaitu sebanyak 1806 rumah tangga yang tinggal di perdesaan tidak terjangkit penyakit malaria. Sedangkan rumah tangga yang tinggal di daerah perkotaan adalah sebanyak 992 rumah tangga, dengan 342 rumah tangga diantaranya (sebesar 10,1 persen dari total rumah tangga keseluruhan) terjangkit penyakit malaria dan sisanya yaitu sebanyak 650 rumah tangga yang tinggal di perkotaan tidak terjangkit penyakit malaria. Sehingga dapat disimpulkan bahwa rumah tangga yang tinggal di daerah

perdesaan lebih rentan terserang penyakit malaria dibandingkan dengan rumah tangga yang tinggal daerah perkotaan.

Tabel 4. 1 Karakteristik Rumah Tangga Berdasarkan Wilayah Tempat Tinggal

		Penyakit Malaria Pada Rumah Tangga		Total
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Wilayah Tempat Tinggal	Perkotaan	342	650	992
		10,1%	19,3%	29,4%
	Perdesaan	575	1806	2381
		17%	53,5%	70,6%
Total		917	2456	3373
		27,2%	72,8%	100%

4.1.3 Jenis Kelamin Kepala Keluarga

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan jenis kelamin kepala keluarga dapat dilihat pada Tabel 4.2.

Tabel 4. 2 Karakteristik Rumah Tangga Berdasarkan Jenis Kelamin Kepala Keluarga

		Penyakit Malaria Pada Rumah Tangga		Total
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Jenis Kelamin	Laki-laki	845	2163	3008
		25,1%	64,1%	89,2%
	Perempuan	72	293	365
		2,1%	8,7%	10,8%
Total		917	2456	3373
		27,2%	72,8%	100%

Berdasarkan Tabel 4.2, dapat diketahui bahwa mayoritas kepala rumah tangga di Provinsi Papua Barat adalah laki-laki yaitu sebanyak 3008 orang, dengan 845 rumah tangga diantaranya terjangkit (sebesar 25,1 persen dari total rumah tangga keseluruhan) penyakit malaria dan sisanya yaitu sebesar 2163 rumah tangga tidak terjangkit malaria. Sedangkan sebanyak 365 rumah tangga mempunyai kepala rumah tangga berjenis kelamin perempuan, dengan 72 rumah tangga diantaranya (sebesar 2,1 persen dari total rumah tangga keseluruhan) terjangkit penyakit malaria dan 293 rumah tangga sisanya tidak terjangkit penyakit malaria. Hal ini menunjukkan bahwa rumah tangga dengan kepala keluarga berjenis kelamin laki-laki rentan terserang penyakit malaria. Hal ini dikarenakan perempuan mempunyai respon imun yang lebih kuat dibandingkan dengan laki-laki (Harijanto, 2000).

4.1.4 Status Kawin

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan status kawin kepala keluarga dapat dilihat pada Tabel 4.3.

Berdasarkan Tabel 4.3, dapat diketahui bahwa mayoritas kepala rumah tangga di Provinsi Papua Barat berstatus menikah yaitu sebanyak 2906 orang, dengan 833 rumah tangga diantaranya (sebesar 24,7 persen dari total rumah tangga) terjangkit penyakit malaria dan sisanya yaitu sebesar 2073 rumah tangga tidak terjangkit malaria. Sedangkan jumlah rumah tangga yang kepala keluarganya berstatus hidup bersama hanya terdiri dari 4 rumah tangga dengan tidak ada rumah tangga diantaranya yang terjangkit malaria. Sehingga dapat disimpulkan bahwa rumah tangga yang terjangkit malaria terbanyak adalah rumah tangga dengan kepala keluarga berstatus menikah.

Tabel 4. 3 Karakteristik Rumah Tangga Berdasarkan Status Kawin Kepala Keluarga

		Penyakit Malaria Pada Rumah Tangga		Total
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Status Kawin KK	Belum menikah	14 0.4%	109 3.2%	123 3.6%
	Menikah	833 24.7%	2073 61.5%	2906 86.2%
	Hidup bersama	0 0%	4 0.1%	4 0.1%
	Cerai hidup	8 0.2%	44 1.3%	52 1.5%
	Hidup terpisah	2 0.1%	28 0.8%	30 0.9%
	Cerai mati	60 1.8%	198 5.9%	258 7.6%
	Total	917 27.2%	2456 72.8%	3373 100%

4.1.5 Tingkat Pendidikan Kepala Keluarga

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan tingkat pendidikan terakhir kepala keluarga dapat dilihat pada Tabel 4.4.

Berdasarkan Tabel 4.4, dapat diketahui bahwa mayoritas rumah tangga di Provinsi Papua Barat mempunyai kepala keluarga dengan pendidikan terakhir tamat SLTA/MA yaitu sebanyak 956 orang, dengan 306 orang diantaranya (sebesar 9,1 persen dari total rumah tangga) terjangkit penyakit malaria dan sisanya yaitu sebanyak 650 rumah tangga tidak terjangkit malaria. Sedangkan rumah tangga yang memiliki kepala keluarga tidak atau belum pernah sekolah, hanya sebanyak 31 rumah tangga yang terjangkit malaria. Tingkat pendidikan sebenarnya bukan sebagai faktor langsung terhadap kejadian malaria karena jika tingkat pendidikan tinggi namun tidak diikuti dengan pemahaman sikap serta perilaku positif dalam pencegahan penyakit malaria

maka akan mempunyai resiko yang sama terkena penyakit malaria.

Tabel 4. 4 Karakteristik Rumah Tangga Berdasarkan Tingkat Pendidikan Kepala Keluarga

	Penyakit Malaria Pada Rumah Tangga			Total
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Pendidikan Terakhir KK	Tidak/belum pernah sekolah	31 0.9%	163 4.8%	194 5.8%
	Tidak tamat SD/MI	93 2.8%	392 11.6%	485 14.4%
	Tamat SD/MI	200 5.9%	604 17.9%	804 23.8%
	Tamat SLTP/MTS	158 4.7%	431 12.8%	589 17.5%
	Tamat SLTA/MA	306 9.1%	650 19.3%	956 28.3%
	Tamat D1/D2/D3	54 1.6%	111 3.3%	165 4.9%
	Tamat PT	75 2.2%	105 3.1%	180 5.3%
	Total	917 27.2%	2456 72.8%	3373 100%

4.1.6 Pekerjaan Kepala Keluarga

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan pekerjaan kepala keluarga dapat dilihat pada Tabel 4.5.

Berdasarkan Tabel 4.5, dapat diketahui bahwa mayoritas rumah tangga di Provinsi Papua Barat mempunyai kepala keluarga dengan pekerjaan sebagai petani yaitu sebanyak 806 orang, dengan 179 orang diantaranya (sebesar 5,3 persen dari total rumah tangga keseluruhan) terjangkau penyakit malaria dan sisanya yaitu sebanyak 627 orang tidak terjangkau penyakit malaria. Pekerjaan kepala keluarga dengan urutan terbanyak kedua adalah wiraswasta, yaitu sebanyak 615 orang, dengan 202

kedua adalah wiraswasta yaitu sebanyak 615 orang, dengan 202 orang diantaranya (sebesar 6 persen dari total rumah tangga keseluruhan) terjangkit penyakit malaria dan sisanya sebanyak 413 orang tidak terjangkit malaria. Hal ini menunjukkan bahwa bekerja sebagai petani memberikan resiko tinggi terjangkit malaria karena bekerja di tempat terbuka dan terdapat genangan air atau rawa-rawa sebagai tempat tempat perindukan nyamuk.

Tabel 4. 5 Karakteristik Rumah Tangga Berdasarkan Pekerjaan Kepala Keluarga

	Penyakit Malaria Pada Rumah Tangga		Total
	Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Pekerjaan KK	PNS/TNI/Polri/BUMD	160 4.7%	261 7.7% 421 12.5%
	Pegawai swasta	85 2.5%	231 6.8% 316 9.4%
	Wiraswasta	202 6.0%	413 12.2% 615 18.2%
	Petani	179 5.3%	627 18.6% 806 23.9%
	Nelayan	74 2.2%	264 7.8% 338 10.0%
	Buruh	50 1.5%	122 3.6% 172 5.1%
	Lainnya	56 1.7%	144 4.3% 200 5.9%
	Tidak bekerja	111 3.3%	394 11.7% 505 15.0%
	Total	917 27.2%	2456 72.8% 3373 100%

4.1.7 Status Ekonomi Rumah Tangga

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan status ekonomi rumah tangga dapat dilihat pada Tabel 4.6.

Tabel 4. 6 Karakteristik Rumah Tangga Berdasarkan Status Ekonomi

Penyakit Malaria Pada Rumah Tangga				
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	Total
Status Ekonomi	Kuintil 1-3	618 18,3%	1922 57%	2540 75,3%
	Kuintil 4-5	299 8,9%	534 15,8%	833 24,7%
	Total	917 27,2%	2456 72,8%	3373 100%

Berdasarkan Tabel 4.6, dapat diketahui bahwa mayoritas rumah tangga di Provinsi Papua Barat memiliki status ekonomi pada kuintil 1-3, dimana maksud dari kuintil 1-3 adalah terbawah, menengah bawah, dan menengah, yaitu sebanyak 2540 rumah tangga, dengan 618 rumah tangga diantaranya (sebesar 18,3 persen dari total rumah tangga) terjangkit malaria dan 1922 rumah tangga sisanya tidak terjangkit malaria. Sedangkan rumah tangga dengan status ekonomi kuintil 4-5 (menengah atas dan teratas) sebanyak 833 rumah tangga, dengan 299 rumah tangga diantaranya (sebesar 8,9 persen dari total rumah tangga) terjangkit malaria dan 2456 rumah tangga sisanya tidak terjangkit malaria. Hal ini menunjukkan bahwa rumah tangga dengan status ekonomi kuintil 1-3 rentan terjangkit penyakit malaria bila dibandingkan dengan rumah tangga berstatus ekonomi kuintil 4-5.

4.1.8 Umur Kepala Keluarga dan Jumlah Pemakaian Air Rumah Tangga

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan umur kepala keluarga dan jumlah pemakaian air dalam sehari dapat dilihat pada Tabel 4.7.

Berdasarkan Tabel 4.7 dapat diperoleh informasi bahwa umur kepala rumah tangga di Provinsi Papua Barat rata-rata 44 tahun dengan sebaran data umur kepala keluarga yaitu 12,74

tahun. Umur kepala keluarga minimal adalah 10 tahun dan maksimal adalah 97 tahun. Sedangkan jumlah pemakaian air untuk keperluan rumah tangga rata-rata 268 liter per hari dengan sebaran data jumlah pemakaian air cukup tinggi yaitu 265,88. Jumlah pemakaian air untuk keperluan rumah tangga minimal 10 liter per hari dan maksimal 4001 liter per hari.

Tabel 4. 7 Karakteristik Rumah Tangga Berdasarkan Umur Kepala Keluarga dan Jumlah Pemakaian Air

Variabel		Rata-rata	St. Dev	Min	Maks
Umur (tahun)	X_3	44.32	12.74	10	97
Jumlah Pemakaian Air (liter/hari)	X_7	268.09	265.88	10	4001

4.1.9 Kebiasaan Mengolah Air Minum Sebelum Dikonsumsi

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan kebiasaan mengolah air minum sebelum dikonsumsi dapat dilihat pada Tabel 4.8.

Tabel 4. 8 Karakteristik Rumah Tangga Berdasarkan Pengolahan Air Minum

		Penyakit Malaria Pada Rumah Tangga		Total
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Melakukan Pengolahan Air Minum Sebelum Dikonsumsi	Ya	697	1976	2673
		20,7%	58,6%	79,2%
	Tidak	220	480	700
		6,5%	14,2%	20,8%
Total		917	2456	3373
		27,2%	72,8%	100%

Berdasarkan Tabel 4.8, diperoleh informasi bahwa, mayoritas rumah tangga di Provinsi Papua Barat melakukan pengolahan air minum sebelum dikonsumsi yaitu sebanyak 2673 rumah tangga, dengan 697 rumah tangga diantaranya (sebesar 20,7 persen dari total rumah tangga) terjangkit penyakit malaria

dan 1976 rumah tangga sisanya tidak terjangkau malaria. Sedangkan rumah tangga yang tidak melakukan pengolahan air minum sebelum dikonsumsi sebanyak 700 rumah tangga, dengan 220 rumah tangga diantaranya (sebesar 6,5 persen dari total rumah tangga) terjangkau penyakit malaria dan 480 rumah tangga sisanya tidak terjangkau penyakit malaria. Hal ini menunjukkan bahwa masyarakat Provinsi Papua Barat sadar akan kebersihan minuman yang dikonsumsi.

4.1.10 Adanya Layanan Kesehatan Gratis

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan adanya pelayanan kesehatan gratis selama satu tahun terakhir dapat dilihat pada Tabel 4.9.

Tabel 4. 9 Karakteristik Rumah Tangga Berdasarkan Adanya Layanan Kesehatan Gratis

		Penyakit Malaria Pada Rumah Tangga		Total
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Layanan Kesehatan Gratis	Ya	435 12,9%	1236 36,6%	1671 49,5%
	Tidak	482 14,3%	1220 36,2%	1702 50,5%
Total		917 27,2%	2456 72,8%	3373 100%

Berdasarkan Tabel 4.9, diperoleh informasi bahwa mayoritas rumah tangga di Provinsi Papua Barat tidak mendapatkan layanan kesehatan gratis dalam satu tahun terakhir selama tahun 2013, yaitu sebanyak 1702 rumah tangga, dengan 482 rumah tangga diantaranya (sebesar 14,3 persen dari total rumah tangga keseluruhan) terjangkau penyakit malaria, sedangkan 1220 rumah tangga sisanya tidak terjangkau penyakit malaria. Sedangkan sebanyak 1671 rumah tangga di Papua Barat mendapatkan layanan kesehatan gratis dengan 435 rumah tangga diantaranya (sebesar 12,9 persen dari total rumah tangga keseluruhan) terjangkau penyakit malaria. Namun jika

dibandingkan antara rumah tangga yang mendapatkan pelayanan kesehatan gratis dengan rumah tangga yang tidak mendapatkan pelayanan kesehatan gratis tidak berbeda jauh jumlahnya.

4.1.11 Jenis Sumber Air Utama Rumah Tangga

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan jenis sumber air utama yang digunakan oleh rumah tangga dapat dilihat pada Tabel 4.10.

Tabel 4. 10 Karakteristik Rumah Tangga Berdasarkan Jenis Sumber Air Utama

		Penyakit Malaria Pada Rumah Tangga		Total
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Jenis sumber air utama	Air tidak terlindungi	106 3.1%	375 11.1%	481 14.3%
	Air terlindungi	811 24%	2081 61.7%	2892 85.7%
	Total	917 27.2%	2456 72.8%	3373 100%

Berdasarkan Tabel 4.10, diperoleh informasi bahwa mayoritas rumah tangga di Papua Barat menggunakan air terlindungi sebagai sumber air utama, yaitu sebanyak 2892 rumah tangga, dengan 811 rumah tangga diantaranya (sebesar 24 persen dari total rumah tangga keseluruhan) terjangkit penyakit malaria dan sisanya yaitu sebanyak 2081 rumah tangga tidak terjangkit penyakit malaria. Sedangkan sebanyak 481 rumah tangga di Provinsi Papua Barat menggunakan air tidak terlindungi sebagai sumber air utama, dengan 106 rumah tangga diantaranya (sebesar 3,1 persen dari total rumah tangga keseluruhan) terjangkit penyakit malaria. Hal ini menunjukkan bahwa masyarakat di Provinsi Papua Barat sudah sadar akan kebersihan air yang digunakan untuk kebutuhan sehari-hari.

4.1.12 Adanya Pencegahan Gigitan Nyamuk

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan adanya upaya pencegahan terhadap gigitan nyamuk dapat dilihat pada Tabel 4.11.

Tabel 4. 11 Karakteristik Rumah Tangga Berdasarkan Adanya Pencegahan Gigitan Nyamuk

		Penyakit Malaria Pada Rumah Tangga		Total
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Adanya pencegahan gigitan nyamuk	Tidak ada pencegahan	149 4.4%	383 11.4%	532 15.8%
	Ada pencegahan gigitan nyamuk	768 22.8%	2073 61.5%	2841 84.2%
	Total	917 27.2%	2456 72.8%	3373 100%

Berdasarkan Tabel 4.11, diperoleh informasi bahwa mayoritas rumah tangga di Provinsi Papua Barat telah melakukan upaya untuk mencegah gigitan nyamuk, yaitu sebanyak 2841 rumah tangga, dengan 768 rumah tangga diantaranya (sebesar 22,8 persen dari total rumah tangga keseluruhan) terjangkit penyakit malaria. Meskipun rumah tangga telah melakukan upaya pencegahan terhadap gigitan nyamuk, namun rumah tangga yang terjangkit malaria masih cukup banyak. Sedangkan sebanyak 532 rumah tangga di Provinsi Papua Barat tidak melakukan upaya pencegahan terhadap gigitan nyamuk dengan 149 rumah tangga diantaranya terjangkit penyakit malaria (sebesar 4,4 persen dari total rumah tangga keseluruhan).

4.1.13 Lingkungan Tempat Tinggal

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan lingkungan tempat tinggal, apakah rumah terletak di daerah kumuh atau tidak dapat dilihat pada Tabel 4.12.

Tabel 4. 12 Karakteristik Rumah Tangga Berdasarkan Lingkungan Tempat Tinggal

		Penyakit Malaria Pada Rumah Tangga		Total
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Daerah Kumuh	Ya	154 4,6%	498 14,8%	652 19,3%
	Tidak	763 22,6%	1958 58%	2721 80,7%
Total		917 27,2%	2456 72,8%	3373 100%

Berdasarkan Tabel 4.12, diperoleh informasi bahwa mayoritas rumah tangga di Provinsi Papua Barat tidak tinggal di daerah kumuh yaitu sebanyak 2721 rumah tangga, dengan 763 rumah tangga diantaranya (sebesar 22,6 persen dari total rumah tangga keseluruhan) terjangkit malaria dan 1958 rumah tangga sisanya tidak terjangkit malaria. Sedangkan rumah tangga yang tinggal di daerah kumuh sebanyak 652 rumah tangga, dengan 154 rumah tangga diantaranya (sebesar 4,6 persen dari total rumah tangga keseluruhan) terjangkit malaria dan 498 rumah tangga sisanya tidak terjangkit malaria. Hal ini menunjukkan bahwa masyarakat Provinsi Papua Barat sudah sadar akan kebersihan lingkungan sekitar.

4.1.14Kepadatan Hunian

Karakteristik rumah tangga di Provinsi Papua Barat berdasarkan kepadatan hunian atau rumah dapat dilihat pada Tabel 4.13.

Berdasarkan Tabel 4.13, diperoleh informasi bahwa mayoritas rumah tangga di Provinsi Papua Barat mempunyai kepadatan hunian yang rendah, yaitu sebanyak 2748 rumah tangga mempunyai ruang gerak didalam rumah lebih dari 8m² setiap orang, dengan 729 rumah tangga diantaranya (sebesar 21,6 persen dari total rumah tangga keseluruhan) terjangkit penyakit malaria. Sedangkan rumah tangga yang tergolong padat hunian

sebanyak 625 rumah tangga dengan 188 rumah tangga diantaranya terjangkit penyakit malaria (sebesar 5,6 persen dari total rumah tangga keseluruhan).

Tabel 4. 13 Karakteristik Rumah Tangga Berdasarkan Kepadatan Hunian

		Penyakit Malaria Pada Rumah Tangga		Total
		Ada ART Terjangkit Malaria	Tidak Ada ART Terjangkit Malaria	
Kepadatan Hunian	< 8 m ² /org (padat)	188 5,6%	437 13%	625 18,5%
	≥8 m ² /org (tidak padat)	729 21,6%	2019 59,9%	2748 81,5%
Total		917 27,2%	2456 72,8%	3373 100%

4. 2 Uji Independensi

Uji independensi dilakukan untuk mengetahui adanya hubungan antara variabel prediktor dengan variabel respon. Pengujian independensi menggunakan uji *Chi-Square* dengan hipotesis uji sebagai berikut:

H_0 : Tidak terdapat hubungan antara variabel prediktor dengan variabel respon

H_1 : Terdapat hubungan antara variabel prediktor dengan variabel respon

Hasil pengujian independensi antara variabel prediktor dengan variabel respon dapat dilihat pada Tabel 4.14.

Tabel 4. 14 Hasil uji independensi

Variabel Prediktor	Sig.
Wilayah tempat tinggal (X_1)	0,000*
Jenis kelamin kepala keluarga (X_2)	0,001*
Status kawin kepala keluarga (X_3)	0,000*
Pendidikan terakhir kepala keluarga (X_5)	0,067
Pekerjaan kepala keluarga (X_6)	0,000*
Status ekonomi (X_7)	0,000*
Pengolahan air minum (X_9)	0,005*
Pelayanan kesehatan gratis (X_{10})	0,135

Tabel 4. 14 Hasil uji independensi (Lanjutan)

Variabel Prediktor	Sig.
Jenis sumber air utama (X_{11})	0,006*
Pencegahan gigitan nyamuk (X_{12})	0,643
Daerah kumuh (X_{13})	0,023*
Kepadatan hunian (X_{14})	0,072

Tabel 4.14 menunjukkan bahwa sebanyak delapan variabel prediktor memiliki nilai signifikansi kurang dari taraf signifikansi (nilai alpha) yang ditentukan sebesar 0,05, sehingga berdasarkan pengujian hipotesis, kedelapan variabel tersebut menolak hipotesis awal, maka dapat dikatakan bahwa kedelapan variabel tersebut memiliki hubungan dengan variabel respon. Berdasarkan pengujian independensi, wilayah tempat tinggal (X_1), jenis kelamin kepala keluarga (X_2), status kawin kepala keluarga (X_3), pekerjaan kepala keluarga (X_6), status ekonomi (X_7), pengolahan air minum (X_9), jenis sumber air utama (X_{11}), dan daerah kumuh (X_{13}) memiliki hubungan yang signifikan dengan penyakit malaria pada rumah tangga di Provinsi Papua Barat.

4.3 Klasifikasi Status Rumah Tangga terhadap Penyakit Malaria di Provinsi Papua Barat dengan Pohon Klasifikasi

Berdasarkan pada tujuan penelitian ini yakni ingin memperoleh informasi yang berkaitan dengan variabel prediktor apa yang menjadi faktor paling penting (dominan) dalam menentukan klasifikasi rumah tangga malaria di Provinsi Papua Barat, maka dilakukan analisis klasifikasi status rumah tangga terhadap penyakit malaria di Provinsi Papua Barat dengan menggunakan pendekatan *classification and regression tree* (CART). Adapun variabel respon yang digunakan dalam penelitian ini berupa data kategorik yaitu rumah tangga dengan anggota rumah tangga terjangkit malaria dan rumah tangga dengan anggota rumah tangga tidak terjangkit malaria, sehingga metode analisis yang digunakan yakni pendekatan CART akan menghasilkan suatu pohon klasifikasi (*classification tree*).

Sesuai dengan prosedur algoritma CART yang telah dijelaskan pada bab tinjauan pustaka, maka tahapan pertama yang

dilakukan adalah pembentukan pohon klasifikasi. Metode pemilihan pemilah pada pembentukan pohon klasifikasi dengan menggunakan data penelitian ini menggunakan metode *test sample* karena jumlah data pada penelitian ini sebanyak 3.373 unit rumah tangga setelah dilakukan pra pemrosesan yaitu *data cleaning*. Setelah dilakukan proses *data cleaning*, jumlah rumah tangga yang anggotanya tidak terjangkit malaria lebih banyak daripada yang terjangkit malaria, yakni sebanyak 2.456 rumah tangga yang anggotanya tidak terjangkit malaria, sedangkan rumah tangga yang anggotanya terjangkit malaria sebanyak 917 rumah tangga. Sehingga perlu dilakukan pra-pemrosesan dengan menggunakan metode SMOTE guna menyeimbangkan jumlah anggota dalam kelas minor, dalam penelitian ini adalah kelas rumah tangga dengan anggota rumah tangga terjangkit malaria. Setelah dilakukan proses SMOTE jumlah data menjadi 4.290 unit rumah tangga, dengan jumlah masing-masing kelas adalah 2.456 rumah tangga untuk kelas rumah tangga dengan anggota rumah tangga tidak terjangkit malaria dan 1.834 rumah tangga untuk kelas rumah tangga dengan anggota rumah tangga terjangkit malaria. Kemudian data tersebut akan dibagi menjadi data *learning* dan *testing* dengan kombinasi proporsi yang dicobakan yaitu 75%:25%, 80%:20%, 85%:15%, 90%:10%, dan 95%:5%. Berikut merupakan hasil dari pengolahan data rumah tangga penderita penyakit malaria di Provinsi Papua Barat dengan mencobakan kombinasi data *learning* dan *testing* serta fungsi keheterogenan Indeks Gini dan Indeks Twoing untuk data sebelum melalui pra-pemrosesan SMOTE dan sesudah pra-pemrosesan SMOTE yang ditampilkan pada Tabel 4.15, agar dapat ditentukan kombinasi dan fungsi keheterogenan pemilihan pembentukan pohon klasifikasi terbaik, yakni yang dapat memberikan hasil ketepatan klasifikasi data *testing* tertinggi atau jumlah simpul terkecil.

Tabel 4.15 Ketepatan Klasifikasi Sebelum dan Sesudah Pra-pemrosesan SMOTE

	Fungsi Ke-heterogen-an	L	Frek. Data	T	Frek. Data	Ketepatan Klasifikasi (%)		Banyak Simpul (Node)
						L	T	
Sebelum SMOTE	Indeks Gini	75%	2530	25%	843	71,4	63,3	65
		80%	2698	20%	675	65,5	64,9	5
		85%	2867	15%	506	62,9	63,6	11
		90%	3036	10%	337	63,6	65,9	9
		95%	3204	5%	169	64,7	65,1	10
	Indeks Twoing	75%	2530	25%	843	71,1	61,1	72
		80%	2698	20%	675	65,5	64,9	5
		85%	2867	15%	506	62,9	63,6	11
		90%	3036	10%	337	63,6	65,9	9
		95%	3204	5%	169	64,7	65,1	10
Sesudah SMOTE	Indeks Gini	75%	3218	25%	1072	82,4	70,7	200
		80%	3432	20%	858	79,3	67,8	138
		85%	3646	15%	644	83	68,5	231
		90%	3861	10%	429	81,5	69,7	205
		95%	4076	5%	214	65,3	68,7	12
	Indeks Twoing	75%	3218	25%	1072	82,4	70,6	200
		80%	3432	20%	858	79,2	67,8	138
		85%	3646	15%	644	83	68,3	231
		90%	3861	10%	429	81,5	69,7	205
		95%	4076	5%	214	65,3	68,7	12

Ket : L = *Learning*, T = *Testing*

Berdasarkan Tabel 4.15, diperoleh informasi bahwa dengan menggunakan kombinasi proporsi 75%:25%, baik dengan menggunakan fungsi keheterogenan Indeks Gini maupun Indeks Twoing pada data sesudah pra-pemrosesan SMOTE diperoleh persentase ketepatan klasifikasi data *testing* yang tertinggi, yakni sebesar 70,7% dengan jumlah simpul terminal pohon klasifikasi sebanyak 200. Ketepatan klasifikasi data *testing* dijadikan sebagai dasar pemilihan metode pembentukan model pohon klasifikasi karena dapat memberikan gambaran kebaikan model pohon klasifikasi yang nantinya terbentuk untuk mengklasikasikan data baru. Meskipun kombinasi data *learning* dan *testing* 75%:25% mempunyai ketepatan klasifikasi tertinggi, namun jumlah simpul terminal yang dihasilkan terlalu banyak sehingga akan membuat pohon klasifikasi menjadi lebar. Berdasarkan prinsip *parsimony*,

maka kombinasi data *learning* dan *testing* yang digunakan adalah 95%:5% karena memiliki simpul terminal paling sederhana, yakni 12 simpul terminal serta selisih persentase ketepatan klasifikasi data *testing* tidak berbeda jauh jika dibandingkan dengan kombinasi data *learning* dan data *testing* 75%:25%. Sehingga akan lebih mudah dalam menginterpretasikan hasil pohon klasifikasi.

4.3.1 Pembentukan Pohon Klasifikasi Maksimal

Pembentukan pohon klasifikasi pada mulanya berupa pohon klasifikasi maksimal, yakni pohon klasifikasi yang memiliki jumlah simpul terbanyak. Setiap pemilah pada suatu simpul hanya bergantung pada nilai yang berasal dari satu variabel prediktor, dimana setiap variabel prediktor mempunyai sejumlah kemungkinan pemilah, tergantung skala data variabel prediktor tersebut.

Dari berbagai kemungkinan pemilah setiap variabel prediktor, selanjutnya dihitung ukuran keheterogenan simpul. Dikarenakan hasil perhitungan dari Indeks Gini dan Twoing memberikan hasil yang sama, maka Indeks Gini akan digunakan untuk menentukan *goodness of split* dari masing-masing pemilah, dengan alasan lebih sederhana. Pemilah yang terpilih adalah variabel pemilah yang memiliki nilai *goodness of split* tertinggi. Pemilah yang terpilih merupakan variabel yang terpenting dalam mengklasifikasikan data pengamatan. Hasil pengolahan data diperoleh urutan kepentingan variabel faktor yang berpengaruh terhadap status rumah tangga terhadap penyakit malaria, seperti ditampilkan pada Tabel 4.16.

Tabel 4. 16 Variabel Penting Pembentukan Pohon Klasifikasi Maksimal

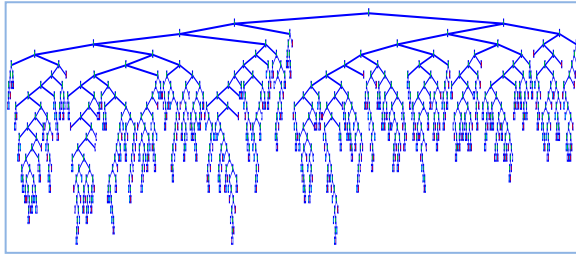
Variabel	Skor	
Umur KK (X_4)	100.00	
Jumlah Pemakaian Air (X_8)	65.81	
Pendidikan KK (X_5)	57.40	
Pekerjaan KK (X_6)	53.93	
Wilayah Tempat Tinggal (X_1)	23.68	
Status Kawin KK (X_3)	17.70	
Daerah Kumuh (X_{13})	17.47	

Tabel 4. 16 Variabel Penting Pembentukan Pohon Klasifikasi Maksimal (Lanjutan)

Variabel	Skor	
Jenis Kelamin KK (X_2)	15.97	
Sumber Air Utama (X_{11})	15.57	
Pengolahan Air Minum (X_9)	15.10	
Kepadatan Hunian (X_{14})	14.03	
Status Ekonomi (X_7)	13.19	
Pencegahan Gigitan Nyamuk (X_{12})	12.26	
Pelayanan Kesehatan Gratis (X_{10})	10.44	

Berdasarkan Tabel 4.17, diperoleh informasi bahwa semua variabel prediktor menjadi pembangun dalam pembentukan pohon klasifikasi. Namun berdasarkan skor variabel yang dihasilkan, dapat diketahui bahwa variabel umur kepala keluarga (X_4) mempunyai skor variabel tertinggi yakni 100. Sehingga dapat dikatakan bahwa variabel prediktor umur kepala keluarga memiliki peranan utama atau faktor terpenting dalam mengklasifikasikan status rumah tangga terhadap penyakit malaria di Provinsi Papua Barat tahun 2013. Selain itu, terdapat beberapa variabel yang juga berpengaruh besar dalam pembentukan pohon klasifikasi yaitu jumlah pemakaian air rumah tangga dalam sehari (X_8) berkontribusi sebesar 65,81, pendidikan kepala keluarga (X_5) berkontribusi sebesar 57,40 dan pekerjaan kepala keluarga (X_6) berkontribusi sebesar 53,93. Sedangkan variabel yang lain memiliki skor di bawah 50.

Hasil penyekatan rekursif biner dari data pengamatan yang digunakan akan menghasilkan pohon klasifikasi yang berukuran relatif besar dan tingkat kedalaman (*depth*) yang tinggi. Pohon klasifikasi tersebut disebut pohon klasifikasi maksimal. Pada Gambar 4.3 ditampilkan hasil pohon klasifikasi maksimal dengan jumlah simpul terminal sebanyak 512 simpul dengan tingkat kedalaman sebanyak 25. Jika digambarkan, pohon klasifikasi maksimal tersebut sangat lebar atau kompleks karena memiliki simpul terminal sangat banyak. Berikut merupakan pohon klasifikasi maksimal.

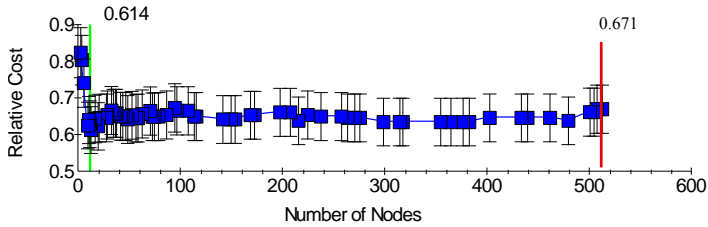


Gambar 4. 3 Topologi Pohon Klasifikasi Maksimal

4.3.2 Pemangkasan Pohon Klasifikasi Maksimal (*Pruning*)

Pohon yang besar dan kompleks dapat memunculkan dugaan adanya kasus *overfitting* (nilai yang dihasilkan melebihi kenyataan yang ada) dan juga dapat mempersulit peneliti dalam hal interpretasi hasil klasifikasi. Untuk memudahkan proses analisis, pohon klasifikasi maksimal yang dihasilkan kemudian dilakukan pemangkasan pohon secara iteratif berdasarkan kriteria tertentu yang telah ditentukan peneliti. Setiap hasil pemangkasan memiliki nilai *relative cost* tertentu, sehingga kemudian dipilih hasil pemangkasan dengan nilai *relative cost* yang minimum.

Pemangkasan pohon dilakukan dengan metode *test sample estimate*. Metode ini lebih sesuai digunakan untuk pemangkasan dalam data pengamatan yang berjumlah besar. Algoritma dari *test sample estimate* ini adalah dengan membagi data penelitian menjadi dua bagian yaitu L_1 untuk data *learning* dan L_2 untuk data *testing*. Data *learning* berguna untuk membentuk pohon T, sedangkan data *testing* berguna dalam pendugaan atau mengestimasi *misclassification rate error* atau disebut juga sebagai *resubstitution estimate* (penduga pengganti). Gambar 4.4 menampilkan adanya perbedaan nilai *relative cost* yang dihasilkan oleh pohon klasifikasi maksimal dengan pohon klasifikasi yang dianggap optimal. Pohon klasifikasi maksimal ditunjukkan oleh garis berwarna merah dan pohon klasifikasi optimal ditunjukkan oleh garis berwarna hijau.

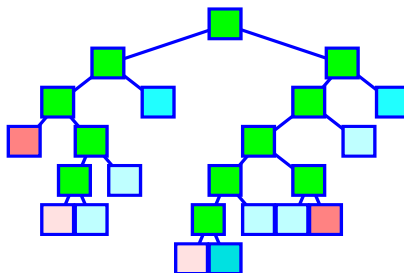


Gambar 4. 4 Plot *Relative Cost* dan Banyaknya Simpul Terminal

Berdasarkan Gambar 4.4, dapat diketahui bahwa nilai *relative cost* yang dihasilkan oleh pohon klasifikasi maksimal dengan jumlah simpul terminal 512 adalah sebesar 0,671 (garis merah). Nilai tersebut lebih besar jika dibandingkan dengan nilai *relative cost* pohon klasifikasi optimal yang memiliki simpul terminal sebanyak 12 simpul yaitu sebesar 0,614 (garis hijau). Nilai kompleksitas pohon klasifikasi optimal sebesar 0,003 dan biaya kesalahannya sebesar $0,614 \pm 0,064$ atau antara 0,55 sampai 0,678. Hasil selengkapnya dapat dilihat di Lampiran D, Output D1. Karena nilai *relative cost* pohon klasifikasi optimal lebih kecil maka pohon klasifikasi optimal dipilih sebagai pohon yang layak untuk pohon klasifikasi rumah tangga dengan malaria di Provinsi Papua Barat pada tahun 2013.

4.3.3 Pemilihan Pohon Klasifikasi Optimal

Hasil pemangkasan pohon maksimal secara iteratif menghasilkan pohon klasifikasi optimal dengan jumlah simpul terminal sebanyak 12 simpul dengan kedalaman pohon sebanyak 7. Gambar 4.5 menampilkan topologi pohon klasifikasi optimal.



Gambar 4. 5 Topologi Pohon Klasifikasi Optimal

Pada Gambar 4.5, terdapat warna simpul dengan degradasi warna antara merah, pink, putih, dan kebiruan. Warna-warna inilah yang menunjukkan jumlah simpul terminal sebanyak 12 simpul. Perbedaan warna tersebut ada kaitannya dengan pemberian label kelas masing-masing simpul terminal. Simpul terminal yang berwarna merah bila suatu simpul diklasifikasikan sebagai rumah tangga dengan minimal ada satu anggota rumah tangga yang terjangkit malaria (kelas 1). Jika warna simpul terminal semakin merah kuat atau pekat, hal itu menunjukkan bahwa rumah tangga anggota simpul tersebut yang termasuk kelas 2 persentasenya mendekati 100 persen. Sedangkan jika warna merah semakin memudar dan menuju warna putih, hal itu menunjukkan bahwa rumah tangga anggota simpul tersebut yang termasuk kelas 1 persentasenya menurun. Sedangkan yang berwarna biru bila suatu simpul diklasifikasikan sebagai rumah tangga dengan anggota rumah tangga tidak terjangkit malaria (kelas 2), jika warna simpul terminal semakin biru kuat atau pekat, hal itu menunjukkan bahwa rumah tangga anggota simpul tersebut yang termasuk kelas 2 persentasenya mendekati 100 persen. Sedangkan jika warna biru semakin memudar dan menuju warna putih, hal itu menunjukkan bahwa rumah tangga anggota simpul tersebut yang termasuk kelas 2 persentasenya menurun. Simpul terminal yang memiliki warna paling kuat atau pekat dilakukan interpretasi karakteristik simpul. Pohon klasifikasi optimal yang lebih jelas dan rinci dengan penjelasan masing-masing pemilah simpul dan banyaknya anggota masing-masing simpul dapat dilihat pada Lampiran E, Output E4.

Berdasarkan topologi pohon klasifikasi optimal, diketahui bahwa pekerjaan kepala keluarga merupakan variabel pemilah yang utama dan paling penting dalam menentukan klasifikasi status rumah tangga terhadap penyakit malaria di Provinsi Papua Barat dengan skor sebesar 100 karena mampu memberikan nilai penurunan keheterogenan tertinggi pada simpul utama. Selain itu ada 11 variabel lain yang juga berkontribusi dalam pembentukan pohon klasifikasi optimal, hasil selengkapnya disajikan dalam Tabel 4.17.

Tabel 4. 17 Variabel Penting Pembentukan Pohon Klasifikasi Optimal

Variabel	Skor	
Pekerjaan KK (X_6)	100.00	
Status Kawin KK (X_3)	78.55	
Jenis Kelamin KK (X_2)	52.74	
Pendidikan (X_5)	48.96	
Pengolahan Air (X_9)	47.91	
Wilayah Tempat Tinggal (X_1)	39.53	
Umur KK (X_4)	30.21	
Jumlah Pemakaian Air (X_8)	23.03	
Daerah Kumuh (X_{13})	14.87	
Pelayanan Kesehatan Gratis (X_{10})	12.75	
Status Ekonomi (X_7)	2.70	
Sumber Air Utama (X_{11})	0.00	
Pencegahan Gigitan Nyamuk (X_{12})	0.00	
Kepadatan Hunian (X_4)	0.00	

Variabel pekerjaan kepala keluarga (X_6) memilah simpul utama (simpul 1) menjadi simpul kiri dan simpul kanan dengan ketentuan pekerjaan kepala keluarga yaitu PNS/TNI/POLRI-/BUMD, pegawai swasta dan wiraswasta (kategori 1, 2 dan 3) akan dipilah menjadi simpul kiri (simpul 2), sedangkan jika pekerjaan kepala keluarga sebagai petani, nelayan, buruh, lainnya dan tidak bekerja (kategori 4, 5, 6, 7 dan 8), akan dipilah menjadi simpul kanan (simpul 6). Diperoleh hasil bahwa ada sebanyak 1845 rumah tangga yang kepala keluarganya bekerja sebagai PNS/TNI/POLRI/BUMD, pegawai swasta dan wiraswasta menjadi anggota simpul kiri (simpul 2) dan sisanya sebanyak 2231 rumah tangga yang kepala keluarganya bekerja sebagai petani, nelayan, buruh, lainnya dan tidak bekerja yang menjadi anggota simpul kanan (simpul 6). Lebih jelasnya dapat dilihat di Lampiran E, Output E2 mengenai informasi pemilahan simpul.

Simpul 2 yang beranggotakan 1845 rumah tangga dengan kepala keluarga yang bekerja sebagai PNS/TNI/POLRI/BUMD, pegawai swasta dan wiraswasta, selanjutnya dipilah menjadi simpul baru kiri dan kanan menurut status kawin kepala keluarga (X_3). Jika status kawin kepala keluarganya menikah, maka akan dipilah menjadi anggota simpul kiri baru (simpul 3). Namun jika status kawin kepala keluarga belum menikah, hidup bersama, cerai hidup, hidup terpisah, dan cerai mati, maka rumah tangga

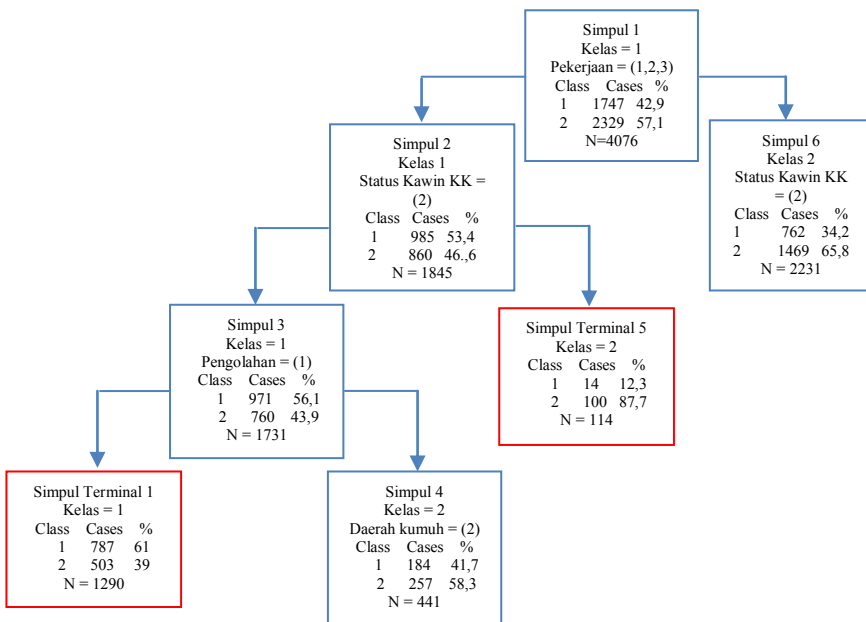
tersebut akan dipilah menjadi simpul kanan baru (simpul terminal 5). Diantara 1845 rumah tangga anggota simpul 2, diperoleh hasil ada sebanyak 1731 rumah tangga menjadi anggota simpul 3 dengan karakteristik kepala keluarga yang bekerja sebagai PNS/TNI/POLRI/BUMD, pegawai swasta dan wiraswasta serta status kawin kepala keluarganya adalah menikah. Sisanya sebanyak 114 rumah tangga yang menjadi anggota simpul terminal 5 dengan karakteristik kepala keluarga yang bekerja sebagai PNS/TNI/POLRI/BUMD, pegawai swasta dan wiraswasta serta status kawin kepala keluarganya adalah belum menikah, hidup bersama, cerai hidup, hidup terpisah, dan cerai mati.

Simpul 3 yang beranggotakan 1731 rumah tangga dengan status kawin kepala keluarganya adalah menikah, selanjutnya dipilah menjadi simpul baru kiri dan kanan menurut kebiasaan mengolah air minum sebelum dikonsumsi. Jika rumah tangga melakukan pengolahan air minum sebelum dikonsumsi, maka rumah tangga tersebut akan dipilah ke simpul baru kiri (simpul terminal 1). Sedangkan jika rumah tangga tidak melakukan pengolahan air minum sebelum dikonsumsi, maka akan dipilah ke simpul kanan baru (simpul 4). Diperoleh hasil bahwa diantara 1731 rumah tangga anggota simpul 3, terdapat 1290 rumah tangga yang menjadi anggota simpul terminal 1 dengan karakteristik kepala keluarga yang bekerja sebagai PNS/TNI/POLRI/BUMD, pegawai swasta dan wiraswasta, status kawin kepala keluarganya adalah menikah dan rumah tangga melakukan pengolahan air minum sebelum dikonsumsi. Sedangkan sisanya sebanyak 441 rumah tangga dipilah ke simpul 4.

Gambar 4.6 merupakan visualisasi potongan struktur pohon klasifikasi optimal untuk penjelasan pemilahan simpul yang dijelaskan di atas, mulai dari pemilahan simpul utama (simpul 1) sampai pemilahan simpul 3 menjadi simpul terminal 1 dan simpul 4. Sehingga diharapkan agar interpretasi struktur pohon klasifikasi yang terbentuk lebih mudah dipahami secara nyata dan jelas. Struktur pohon klasifikasi optimal lebih rinci dengan penjelasan masing-masing pemilah simpul, banyaknya anggota

dan informasi label kelas masing-masing simpul dapat dilihat pada Lampiran E, Output E4.

Suatu simpul akan terus dipilah menjadi simpul anak baru (kiri dan kanan) sesuai prosedur *binary recursive partitioning*, sampai simpul tersebut telah dianggap memiliki anggota yang homogen atau jika simpul tersebut hanya memiliki 1 anggota pengamatan (dalam hal ini 1 rumah tangga) maka simpul akan menjadi simpul terminal dan tidak akan dipilah lagi. Pohon klasifikasi optimal yang terbentuk terdiri atas 12 simpul terminal seperti yang ditunjukkan pada Gambar 4.6. Masing-masing simpul terminal tersebut memiliki karakteristik tertentu dan diprediksi sebagai kelas variabel respon tertentu sesuai dengan label kelas yang diberikan. Berdasarkan hasil penelusuran 12 simpul terminal pohon klasifikasi optimal tersebut, Tabel 4.18 memberikan rangkuman pengklasifikasian kelas rumah tangga menurut indikasi kesamaan label kelas setiap simpul terminal.



Gambar 4. 6 Potongan Struktur Pohon Klasifikasi Optimal

Tabel 4. 18 Kelas Rumah Tangga pada Masing-masing Simpul Terminal

Kelas	Simpul Terminal Ke-	Persentase	Simpul Terminal Ke-	Persentase
(1) Rumah Tangga Terjangkit Penyakit Malaria (4 simpul terminal)	1	61	6	52,8
	2	51,7	10	61
(2) Rumah Tangga Tidak Terjangkit Penyakit Malaria (8 simpul terminal)	3	77,3	8	68
	4	77,6	9	73,6
	5	87,7	11	69,2
	7	90	12	79,4

Penelusuran struktur pohon klasifikasi optimal terhadap simpul terminal dapat memberikan informasi tentang karakteristik kelas simpul terminal dengan persentase tertinggi untuk masing-masing kelas. Berikut adalah karakteristik masing-masing kelas yang disajikan pada Tabel 4.19.

Tabel 4. 19 Karakteristik Kelas Rumah Tangga Menurut Persentase Kelas Tertinggi Simpul Terminal

Kelas	Karakteristik
(1) Rumah Tangga dengan ART terjangkau malaria	Umur kepala keluarga ≥ 65 tahun, jumlah pemakaian air dalam sehari $\geq 214,5$ liter, rumah tangga mendapatkan layanan kesehatan gratis dalam 1 tahun terakhir, status kawin kepala keluarga adalah menikah, serta pekerjaan kepala keluarga sebagai petani, nelayan, buruh, lainnya dan tidak bekerja
(2) Rumah Tangga dengan ART tidak terjangkau malaria	Jenis kelamin kepala keluarga perempuan, pendidikan terakhir kepala keluarga adalah tamat SD/MI, tamat SLTP/MTS dan tamat SLTA/MA, jumlah pemakaian air rumah tangga dalam sehari $\leq 214,5$, rumah tangga mendapatkan layanan kesehatan gratis dalam 1 tahun terakhir, status kawin kepala keluarga adalah menikah serta pekerjaan kepala keluarga sebagai petani, nelayan, buruh, lainnya dan tidak bekerja

4.3.4 Hasil Ketepatan Klasifikasi Pohon Klasifikasi CART

Tingkat keakuratan hasil klasifikasi pohon optimal yang dihasilkan dari data *learning* dapat dihitung berdasarkan Tabel 4.20.

Tabel 4. 20 Klasifikasi Data *Learning* oleh Pohon Klasifikasi Optimal

Kelas Aktual	Kelas Prediksi		Total
	Rumah Tangga Terjangkit Malaria	Rumah Tangga Tidak Terjangkit Malaria	
Rumah Tangga Terjangkit Malaria	1229	518	1747
Rumah Tangga Tidak Terjangkit Malaria	898	1431	2329
Total	2127	1949	4076

Berdasarkan Tabel 4.20, terjadi kesalahan prediksi klasifikasi kelas pengamatan pada masing-masing kelas, yaitu sebanyak 518 rumah tangga yang secara aktual termasuk kelas 1 (terjangkit malaria) namun salah diklasifikasikan sebagai sebagai kelas 2 (tidak terjangkit malaria). Kemudian sebanyak 898 rumah tangga yang secara aktual termasuk kelas 2 (tidak terjangkit malaria) namun salah diklasifikasikan sebagai sebagai kelas 1 (terjangkit malaria).

Berdasarkan Tabel 4.20, diperoleh hasil perhitungan ketepatan klasifikasi data *learning* sebesar 65,3 persen. Artinya bahwa pohon klasifikasi optimal mampu mengklasifikasikan suatu rumah tangga kedalam kelas status terjangkit penyakit malaria dengan tepat sebesar 65,3 persen. Nilai *sensitivity* sebesar 70,3 persen dan nilai *specificity*-nya sebesar 61,4 persen.

Pohon klasifikasi optimal yang terbentuk perlu divalidasi untuk mengetahui apakah pohon klasifikasi tersebut layak dan dapat digunakan untuk mengklasifikasi data baru. Tingkat keakuratan hasil klasifikasi pohon optimal yang dihasilkan dari data *testing* dapat dihitung berdasarkan Tabel 4.21.

Tabel 4. 21 Klasifikasi Data *Testing* oleh Pohon Klasifikasi Optimal

Kelas Aktual	Kelas Prediksi		Total
	Rumah Tangga Terjangkit Malaria	Rumah Tangga Tidak Terjangkit Malaria	
Rumah Tangga Terjangkit Malaria	63	24	87
Rumah Tangga Tidak Terjangkit Malaria	43	84	127
Total	106	108	214

Berdasarkan Tabel 4.21, didapatkan informasi bahwa terjadi kesalahan prediksi klasifikasi kelas pengamatan pada masing-masing kelas, yaitu sebanyak 24 rumah tangga yang secara aktual termasuk kelas 1 (terjangkit malaria) namun salah diklasifikasikan sebagai sebagai kelas 2 (tidak terjangkit malaria). Kemudian sebanyak 43 rumah tangga yang secara aktual termasuk kelas 2 (tidak terjangkit malaria) namun salah diklasifikasikan sebagai sebagai kelas 1 (terjangkit malaria).

Berdasarkan Tabel 4.21, diperoleh hasil perhitungan ketepatan klasifikasi sebesar 68,7 persen. Artinya bahwa pohon klasifikasi optimal mampu mengklasifikasikan suatu rumah tangga kedalam kelas status terjangkit penyakit malaria dengan tepat sebesar 68,7 persen. Nilai *sensitivity* sebesar 72,4 persen dan nilai *specifity*-nya sebesar 66,1 persen.

Berikut adalah perbandingan hasil ketepatan klasifikasi pohon maksimal dengan pohon optimal yang ditunjukkan oleh Tabel 4.22.

Tabel 4. 22 Perbandingan Ketepatan Klasifikasi Pohon Maksimal dan Pohon Optimal

Pohon Klasifikasi	Ketepatan Klasifikasi (%)	
	<i>Learning</i>	<i>Testing</i>
Pohon Maksimal	87,4	66,8
Pohon Optimal	65,3	68,7

Berdasarkan Tabel 4.22, dapat diketahui bahwa untuk data *learning*, ketepatan klasifikasi pohon maksimal lebih tinggi daripada pohon optimal, yakni sebesar 87,4 persen. Hal ini dikarenakan pohon klasifikasi maksimal memiliki simpul yang paling banyak dengan melibatkan lebih banyak variabel prediktor sebagai pemilah simpul sehingga kemungkinan mengklasifikasikan data dengan tepat akan cenderung lebih besar. Sedangkan untuk data *testing*, ketepatan klasifikasi pohon optimal lebih tinggi dibandingkan pohon maksimal, yakni sebesar 68,7 persen.

(Halaman Ini Sengaja Dikosongkan)

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan uraian dan pembahasan yang telah dilakukan, diperoleh kesimpulan bahwa diantara 3.836 rumah tangga di Provinsi Papua Barat tahun 2013, sebanyak 1019 rumah tangga (27 persen) terjangkit penyakit malaria, sedangkan sebanyak 2817 rumah tangga (73 persen) tidak terjangkit penyakit malaria. Ketepatan klasifikasi dengan menggunakan pra-pemrosesan SMOTE menghasilkan tingkat akurasi yang lebih tinggi dibandingkan dengan yang tidak menggunakan pra-pemrosesan SMOTE. Sehingga dalam penelitian ini diputuskan untuk menggunakan pra-pemrosesan SMOTE. Pohon klasifikasi yang layak untuk mengklasifikasikan status rumah tangga penderita malaria di Provinsi Papua Barat yaitu,

- a. Model klasifikasi pohon optimal menggunakan kombinasi data *learning* dan data *testing* sebesar 95%:5% yang memiliki 12 simpul terminal dengan kedalaman pohon sebanyak 8. Data sampel *learning* secara keseluruhan tepat diklasifikasikan oleh pohon klasifikasi sebesar 65,3 persen dan akurasi prediksi data *testing* sebesar 68,7 persen.
- b. Variabel terpenting dalam menentukan status rumah tangga dengan malaria dalam penelitian ini yaitu pekerjaan kepala keluarga dengan skor variabel sebesar 100.

Karakteristik rumah tangga yang terjangkit penyakit malaria menurut penelusuran hasil simpul terminal pohon klasifikasi dengan persentase tertinggi yakni umur kepala keluarga ≥ 65 tahun, jumlah pemakaian air rumah tangga dalam sehari $\geq 214,5$ liter, rumah tangga mendapatkan layanan kesehatan gratis dalam 1 tahun terakhir, status kawin kepala keluarga adalah menikah, serta pekerjaan kepala keluarganya sebagai petani, nelayan, buruh, lainnya dan tidak bekerja.

5.2 Saran

Sebagai saran untuk penelitian berikutnya adalah data yang akan dianalisis perlu disiapkan dengan benar apalagi jika data tersebut jumlahnya sangat besar dan memuat banyak variabel dengan skala campuran agar hasil-hasil yang mungkin tidak logis bisa dihindari dan ketepatan klasifikasi yang dihasilkan lebih tinggi atau lebih baik. Untuk mendapatkan nilai ketepatan klasifikasi yang lebih tinggi pada kasus kelas *imbalanced*, mungkin bisa dicobakan alternatif metode untuk pra-pemrosesan selain metode SMOTE, agar diperoleh nilai akurasi yang lebih tinggi dan jumlah simpul terminal yang lebih sederhana untuk data status rumah tangga terhadap penyakit malaria di Provinsi Papua Barat tahun 2013.

DAFTAR PUSTAKA

- Achmadi, U. F. (2008). *Manajemen Penyakit Berbasis Wilayah*. Jakarta: Universitas Indonesia Press.
- Agresti, A. (2002). *Categorical Data Analysis*. United States of America: John Wiley & Sons.
- Breiman, L., Friedman, J. H., Olshen, R. A., dan Stone, C. J. (1993). *Classification and Regression Trees*. New York: Chapman And Hall.
- Chawla, V. N., Bowyer, K. W., Hall, L. O., dan Kegelmeyer, W. P. (2002). SMOTE : Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, Vol. 16, Hal. 321-357.
- Ekayani, G. (2011). *Analisis Regresi Logistik Biner untuk Mengetahui Faktor-Faktor yang Mempengaruhi Penyakit Malaria di Provinsi Papua Barat*. Tugas Akhir Statistika Institut Teknologi Sepuluh Nopember.
- Frits, W. (2003). *Hubungan Kondisi Fisik Bangunan Rumah dan Tempat Perindukan Nyamuk dengan Kejadian Malaria pada Anak Umur 6-59 bulan di Unit Pelayan Kesehatan di Distrik Fakfak Tahun 2003*. Depok: Universitas Indonesia.
- Harijanto. (2000). *Malaria Epidemiologi Patogenesis Manifestasi Klinis dan Penanganan*. Jakarta: Penerbit Buku Kedokteran EGC.
- Irawan, Y. (2014). *Pengklasifikasian Status HIV/ AIDS Pada Penderita HIV/AIDS di Lembaga Swadaya Masyarakat Orbit Kota Surabaya Menggunakan Metode Classification and Regression Tree*. Tugas Akhir Statistika Institut Teknologi Sepuluh Nopember.

- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis Sixt Edition*. United State of America: Pearson Education, Inc.
- Kementerian Kesehatan RI. (2011). Epidemiologi Malaria di Indonesia. *Buletin Jendela Data dan Informasi Kesehatan*.
- _____. (2013). *Riset Kesehatan Dasar : Riskesdas 2013*. Jakarta: Badan Penelitian dan Pengembangan Kesehatan .
- _____. (2014). Situasi Malaria di Indonesia. *InfoDatin*.
- Lestari, A. S. (2014). *Faktor-Faktor Yang Mempengaruhi Penyakit Malaria Pada Ibu Hamil Di Nusa Tenggara Barat, Nusa Tenggara Timur, Maluku, Maluku Utara, Papua, dan Papua Barat*. Tugas Akhir Statistika Institut Teknologi Sepuluh Nopember.
- Lewis, M. D., & Roger, J. (2000). *An Introduction To Classification And Regression Trees (CART) Analysis*. Presented the 2000.
- Purwanto, A. D. (2009). *Klasifikasi Rumah Tangga Menurut Tempat Berbelanja Barang Kebutuhan Sandang di Kota Yogyakarta Menggunakan Metode Regresi Logistik dan Metode CART*. Tesis Statistika Institut Teknologi Sepuluh Nopember.
- Saikh, A., Budianto, A., dan Yuliani, R. C. (2009). Faktor-faktor Resiko Lingkungan dan Perilaku yang Mempengaruhi Kejadian Kesakitan Malaria di Propinsi Sumatera Selatan (Analisis Lanjut Data Riskesdas 2007). *Jurnal Pembangunan Manusia*.
- Seftiana, D. (2014). *Klasifikasi Rumah Tangga Sangat Miskin di Kabupaten Jombang Menurut Paket Bantuan Rumah*

Tangga yang Diharapkan dengan Pendekatan RF-CART.
Tugas Akhir Statistika Institut Teknologi Sepuluh
Nopember.

- Soemirat, J. (2000). *Epidemiologi Lingkungan*. Yogyakarta: Gadjah Mada University Press.
- Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical Data Analysis Using The SAS System* (Second Edition ed.). Cary, North Carolina: SAS Institute Inc.
- Susilowati, P. A. (2014). *Analisis Regresi Pada Prevalensi Malaria Di Provinsi Maluku Utara, Maluku, Papua Barat Dan Papua Dengan Faktor Yang Mempengaruhinya*. Tugas Akhir Statistika Institut Teknologi Sepuluh Nopember, Institut Teknologi Sepuluh Nopember, Statistics.
- Sutton, C. D. (2005). Classification and Regression Trees, Bagging, and Boosting. In *Handbook of Statistics* (pp. 24, 303-329).
- Trapsilasiwi, R. K. (2014). *Klasifikasi Multiclass untuk Imbalanced Data Menggunakan SMOTE Least Square Support Vector Machine*. Tugas Akhir Statistika Institut Teknologi Sepuluh Nopember.
- Walpole, R. E. (1995). *Pengantar Statistika*. Jakarta: Gramedia Pustaka Utama.

(Halaman Ini Sengaja Dikosongkan)

DAFTAR LAMPIRAN

	Halaman
Lampiran A. Data Penelitian	79
Lampiran B. Uji Independensi	80
Lampiran C. Ketepatan Klasifikasi	86
Output C1. Ketepatan Klasifikasi 75%:25% dengan Indeks Gini.....	86
Output C2. Ketepatan Klasifikasi 75%:25% dengan Indeks Twoing	87
Output C3. Ketepatan Klasifikasi 80%:20% dengan Indeks Gini.....	88
Output C4. Ketepatan Klasifikasi 80%:20% dengan Indeks Twoing	89
Output C5. Ketepatan Klasifikasi 85%:15% dengan Indeks Gini.....	90
Output C6. Ketepatan Klasifikasi 85%:15% dengan Indeks Twoing	92
Output C7. Ketepatan Klasifikasi 90%:10% dengan Indeks Gini.....	93
Output C8. Ketepatan Klasifikasi 90%:10% dengan Indeks Twoing	94
Output C9. Ketepatan Klasifikasi 95%:5% dengan Indeks Gini.....	95
Output C10. Ketepatan Klasifikasi 95%:5% dengan Indeks Twoing	96
Lampiran D. Pohon Klasifikasi Maksimal	98
Output D1. Informasi Pembentukan Pohon Klasifikasi dengan <i>Test Sample</i> Kombinasi Data <i>Learning</i> dan Data <i>Testing</i> 95%:5%	98

Output D2.	Informasi Pemangkasan Pohon Klasifikasi dengan <i>Test Sample</i> Kombinasi Data <i>Learning</i> dan Data <i>Testing</i> 95%:5%	99
Output D3.	Ketepatan Klasifikasi Data <i>Learning</i> dan Data <i>Testing</i> Pohon Klasifikasi Maksimal	100
Lampiran E.	Pohon Klasifikasi Optimal	101
Output E1.	Ketepatan Klasifikasi Data <i>Learning</i> dan Data <i>Testing</i> Pohon Klasifikasi Optimal.....	101
Output E2.	Informasi Pemilahan Simpul Pohon Klasifikasi Optimal.....	101
Output E3.	Informasi Proporsi Setiap Kelas Pada Setiap Simpul Pohon Klasifikasi Optimal.....	108
Output E4.	Struktur Pohon Klasifikasi Optimal	110
Output E6.	Penjelasan Pohon Klasifikasi Optimal	112
Lampiran F.	Surat Keterangan Data.....	114

Lampiran A. Data Penelitian

No.	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄
1	1	2	1	2	32	6	1	1	100	1	2	2	2	2	2
2	1	2	1	2	30	5	7	1	100	1	1	2	2	2	2
3	1	2	1	2	33	2	5	1	100	1	1	2	2	2	2
.
.
.
3836	2	1	1	2	44	4	3	2	120	1	1	2	1	1	1

Y Status rumah tangga terhadap penyakit malaria
X₁ Wilayah tempat tinggal
X₂ Jenis kelamin KK
X₃ Status Kawin KK
X₄ Umur KK
X₅ Pendidikan KK
X₆ Pekerjaan KK
X₇ Status Ekonomi

X₈ Jumlah pemakaian air dalam sehari
X₉ Pengolahan air sebelum dikonsumsi
X₁₀ Layanan kesehatan gratis
X₁₁ Sumber air utama
X₁₂ Pencegahan Gigitan Nyamuk
X₁₃ Daerah kumuh
X₁₄ Kepadatan Hunian

Lampiran B. Uji Independensi**Output B1. Wilayah Tempat Tinggal****Chi-Square Tests**

	Value	Df	Asymp. Sig. (2- sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	37.721 ^a	1	.000		
Continuity Correction ^b	37.201	1	.000		
Likelihood Ratio	36.720	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	37.709	1	.000		
N of Valid Cases	3373				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 269.69.

b. Computed only for a 2x2 table

Output B2. Jenis Kelamin Kepala Keluarga**Chi-Square Tests**

	Value	df	Asymp. Sig. (2- sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	11.508 ^a	1	.001		
Continuity Correction ^b	11.089	1	.001		
Likelihood Ratio	12.205	1	.000		
Fisher's Exact Test				.001	.000
Linear-by-Linear Association	11.505	1	.001		
N of Valid Cases	3373				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 99.23.

b. Computed only for a 2x2 table

Output B3. Status Kawin Kepala Keluarga**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	32.276 ^a	5	.000
Likelihood Ratio	38.625	5	.000
Linear-by-Linear Association	3.241	1	.072
N of Valid Cases	3373		

a. 2 cells (16.7%) have expected count less than 5. The minimum expected count is 1.09.

Output B4. Pendidikan Terakhir Kepala Keluarga**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	63.095 ^a	6	.000
Likelihood Ratio	63.674	6	.000
Linear-by-Linear Association	61.102	1	.000
N of Valid Cases	3373		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 44.86.

Output B5. Pekerjaan Kepala Keluarga**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	57.016 ^a	7	.000
Likelihood Ratio	55.943	7	.000
Linear-by-Linear Association	24.388	1	.000
N of Valid Cases	3373		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 46.76.

Output B6. Status Ekonomi**Chi-Square Tests**

	Value	Df	Asymp. Sig. (2- sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	42.373 ^a	1	.000		
Continuity Correction ^b	41.791	1	.000		
Likelihood Ratio	40.840	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	42.360	1	.000		
N of Valid Cases	3373				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 226.46.

b. Computed only for a 2x2 table

Output B7. Pengolahan Air Minum**Chi-Square Tests**

	Value	Df	Asymp. Sig. (2- sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.030 ^a	1	.005		
Continuity Correction ^b	7.762	1	.005		
Likelihood Ratio	7.855	1	.005		
Fisher's Exact Test				.006	.003
Linear-by-Linear Association	8.028	1	.005		
N of Valid Cases	3373				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 190.31.

b. Computed only for a 2x2 table

Output B8. Layanan Kesehatan Gratis**Chi-Square Tests**

	Value	df	Asymp. Sig. (2- sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.228 ^a	1	.135		
Continuity Correction ^b	2.114	1	.146		
Likelihood Ratio	2.229	1	.135		
Fisher's Exact Test				.141	.073
Linear-by-Linear Association	2.228	1	.136		
N of Valid Cases	3373				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 454.29.

b. Computed only for a 2x2 table

Output B9. Jenis Sumber Air Utama**Chi-Square Tests**

	Value	Df	Asymp. Sig. (2- sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	7.514 ^a	1	.006		
Continuity Correction ^b	7.213	1	.007		
Likelihood Ratio	7.795	1	.005		
Fisher's Exact Test				.006	.003
Linear-by-Linear Association	7.511	1	.006		
N of Valid Cases	3373				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 130.77.

b. Computed only for a 2x2 table

Output B10. Pencegahan Gigitan Nyamuk**Chi-Square Tests**

	Value	Df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.215 ^a	1	.643		
Continuity Correction ^b	.169	1	.681		
Likelihood Ratio	.214	1	.644		
Fisher's Exact Test				.633	.339
Linear-by-Linear Association	.215	1	.643		
N of Valid Cases	3373				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 144.63.

b. Computed only for a 2x2 table

Output B11. Daerah Kumuh**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5.194 ^a	1	.023		
Continuity Correction ^b	4.973	1	.026		
Likelihood Ratio	5.313	1	.021		
Fisher's Exact Test				.024	.012
Linear-by-Linear Association	5.193	1	.023		
N of Valid Cases	3373				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 177.26.

b. Computed only for a 2x2 table

Output B12. Kepadatan Hunian**Chi-Square Tests**

	Value	df	Asymp. Sig. (2- sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.245 ^a	1	.072		
Continuity Correction ^b	3.068	1	.080		
Likelihood Ratio	3.193	1	.074		
Fisher's Exact Test				.073	.041
Linear-by-Linear Association	3.244	1	.072		
N of Valid Cases	3373				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 169.92.

b. Computed only for a 2x2 table

Lampiran C. Ketepatan Klasifikasi**Output C1. Ketepatan Klasifikasi 75%:25% dengan Indeks Gini**

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	325.00	139.00	464.00
2	175.00	433.00	608.00
PRED. TOT.	500.00	572.00	1072.00
CORRECT	0.700	0.712	
SUCCESS IND.	0.268	0.145	
TOT. CORRECT	0.707		

SENSITIVITY: 0.700 SPECIFICITY: 0.712
 FALSE REFERENCE: 0.350 FALSE RESPONSE: 0.243
 REFERENCE = "1", RESPONSE = "2"

 =====
 TEST SAMPLE CLASSIFICATION PROBABILITY TABLE
 =====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.700	0.300	1.000
2	0.288	0.712	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	1132.00	238.00	1370.00
2	329.00	1519.00	1848.00
PRED. TOT.	1461.00	1757.00	3218.00
CORRECT	0.826	0.822	
SUCCESS IND.	0.401	0.248	
TOT. CORRECT	0.824		

SENSITIVITY: 0.826 SPECIFICITY: 0.822
 FALSE REFERENCE: 0.225 FALSE RESPONSE: 0.135
 REFERENCE = "1", RESPONSE = "2"

```
=====
```

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

```
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.826	0.174	1.000
2	0.178	0.822	1.000

```
-----
```

Output C2. Ketepatan Klasifikasi 75%:25% dengan Indeks Twoing

TEST SAMPLE CLASSIFICATION TABLE

```
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	324.00	140.00	464.00
2	175.00	433.00	608.00
PRED. TOT.	499.00	573.00	1072.00
CORRECT	0.698	0.712	
SUCCESS IND.	0.265	0.145	
TOT. CORRECT	0.706		

SENSITIVITY: 0.698 SPECIFICITY: 0.712
 FALSE REFERENCE: 0.351 FALSE RESPONSE: 0.244
 REFERENCE = "1", RESPONSE = "2"

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

```
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.698	0.302	1.000
2	0.288	0.712	1.000

```
-----
```

LEARNING SAMPLE CLASSIFICATION TABLE

```
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	1132.00	238.00	1370.00
2	329.00	1519.00	1848.00
PRED. TOT.	1461.00	1757.00	3218.00

CORRECT	0.826	0.822	
SUCCESS IND.	0.401	0.248	
TOT. CORRECT	0.824		
SENSITIVITY:	0.826	SPECIFICITY:	0.822
FALSE REFERENCE:	0.225	FALSE RESPONSE:	0.135
REFERENCE = "1", RESPONSE = "2"			

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.826	0.174	1.000
2	0.178	0.822	1.000

Output C3. Ketepatan Klasifikasi 80%:20% dengan Indeks Gini

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	235.00	122.00	357.00
2	154.00	347.00	501.00
PRED. TOT.	389.00	469.00	858.00
CORRECT	0.658	0.693	
SUCCESS IND.	0.242	0.109	
TOT. CORRECT	0.678		

SENSITIVITY:	0.658	SPECIFICITY:	0.693
FALSE REFERENCE:	0.396	FALSE RESPONSE:	0.260
REFERENCE = "1", RESPONSE = "2"			

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.658	0.342	1.000
2	0.307	0.693	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	1159.00	318.00	1477.00
2	394.00	1561.00	1955.00
PRED. TOT.	1553.00	1879.00	3432.00
CORRECT	0.785	0.798	
SUCCESS IND.	0.354	0.229	
TOT. CORRECT	0.793		

SENSITIVITY: 0.785 SPECIFICITY: 0.798
 FALSE REFERENCE: 0.254 FALSE RESPONSE: 0.169
 REFERENCE = "1", RESPONSE = "2"

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.785	0.215	1.000
2	0.202	0.798	1.000

Output C4. Ketepatan Klasifikasi 80%:20% dengan Indeks Twoing

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	235.00	122.00	357.00
2	154.00	347.00	501.00
PRED. TOT.	389.00	469.00	858.00
CORRECT	0.658	0.693	
SUCCESS IND.	0.242	0.109	
TOT. CORRECT	0.678		

SENSITIVITY: 0.658 SPECIFICITY: 0.693
 FALSE REFERENCE: 0.396 FALSE RESPONSE: 0.260
 REFERENCE = "1", RESPONSE = "2"

=====			
TEST SAMPLE CLASSIFICATION PROBABILITY TABLE			
=====			
Actual	Predicted Class		Actual
Class	1	2	Total

1	0.658	0.342	1.000
2	0.307	0.693	1.000

=====			
LEARNING SAMPLE CLASSIFICATION TABLE			
=====			
Actual	Predicted Class		Actual
Class	1	2	Total

1	1161.00	316.00	1477.00
2	397.00	1558.00	1955.00

PRED. TOT.	1558.00	1874.00	3432.00
CORRECT	0.786	0.797	
SUCCESS IND.	0.356	0.227	
TOT. CORRECT	0.792		
SENSITIVITY:	0.786	SPECIFICITY:	0.797
FALSE REFERENCE:	0.255	FALSE RESPONSE:	0.169
REFERENCE = "1", RESPONSE = "2"			

=====			
LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE			
=====			
Actual	Predicted Class		Actual
Class	1	2	Total

1	0.786	0.214	1.000
2	0.203	0.797	1.000

Output C5. Ketepatan Klasifikasi 85%:15% dengan Indeks Gini

=====			
TEST SAMPLE CLASSIFICATION TABLE			
=====			
Actual	Predicted Class		Actual
Class	1	2	Total

1	158.00	94.00	252.00
2	109.00	283.00	392.00

PRED. TOT.	267.00	377.00	644.00
CORRECT	0.627	0.722	

SUCCESS IND. 0.236 0.113
 TOT. CORRECT 0.685

SENSITIVITY: 0.627 SPECIFICITY: 0.722
 FALSE REFERENCE: 0.408 FALSE RESPONSE: 0.249
 REFERENCE = "1", RESPONSE = "2"

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.627	0.373	1.000
2	0.278	0.722	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	1310.00	272.00	1582.00
2	347.00	1717.00	2064.00
PRED. TOT.	1657.00	1989.00	3646.00
CORRECT	0.828	0.832	
SUCCESS IND.	0.394	0.266	
TOT. CORRECT	0.830		

SENSITIVITY: 0.828 SPECIFICITY: 0.832
 FALSE REFERENCE: 0.209 FALSE RESPONSE: 0.137
 REFERENCE = "1", RESPONSE = "2"

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.828	0.172	1.000
2	0.168	0.832	1.000

Output C6. Ketepatan Klasifikasi 85%:15% dengan Indeks Twoing

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	158.00	94.00	252.00
2	109.00	283.00	392.00
PRED. TOT.	267.00	377.00	644.00
CORRECT	0.627	0.722	
SUCCESS IND.	0.236	0.113	
TOT. CORRECT	0.685		

SENSITIVITY: 0.627 SPECIFICITY: 0.722
 FALSE REFERENCE: 0.408 FALSE RESPONSE: 0.249
 REFERENCE = "1", RESPONSE = "2"

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.627	0.373	1.000
2	0.278	0.722	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	1310.00	272.00	1582.00
2	347.00	1717.00	2064.00
PRED. TOT.	1657.00	1989.00	3646.00
CORRECT	0.828	0.832	
SUCCESS IND.	0.394	0.266	
TOT. CORRECT	0.830		

SENSITIVITY: 0.828 SPECIFICITY: 0.832
 FALSE REFERENCE: 0.209 FALSE RESPONSE: 0.137
 REFERENCE = "1", RESPONSE = "2"

```
=====
LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.828	0.172	1.000
2	0.168	0.832	1.000

```
-----
```

Output C7. Ketepatan Klasifikasi 90%:10% dengan Indeks Gini

```
=====
TEST SAMPLE CLASSIFICATION TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	137.00	49.00	186.00
2	81.00	162.00	243.00
PRED. TOT.	218.00	211.00	429.00
CORRECT	0.737	0.667	
SUCCESS IND.	0.303	0.100	
TOT. CORRECT	0.697		

```
-----
```

SENSITIVITY: 0.737 SPECIFICITY: 0.667
 FALSE REFERENCE: 0.372 FALSE RESPONSE: 0.232
 REFERENCE = "1", RESPONSE = "2"

```
-----
```

```
=====
TEST SAMPLE CLASSIFICATION PROBABILITY TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.737	0.263	1.000
2	0.333	0.667	1.000

```
-----
```

```
=====
LEARNING SAMPLE CLASSIFICATION TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	1	2	
1	1322.00	326.00	1648.00
2	389.00	1824.00	2213.00
PRED. TOT.	1711.00	2150.00	3861.00
CORRECT	0.802	0.824	

```

SUCCESS IND.          0.375          0.251
TOT. CORRECT          0.815

      SENSITIVITY:          0.802      SPECIFICITY:          0.824
FALSE REFERENCE:          0.227  FALSE RESPONSE:          0.152
REFERENCE = "1", RESPONSE = "2"
-----

```

```

=====
LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE
=====

```

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.802	0.198	1.000
2	0.176	0.824	1.000

Output C8. Ketepatan Klasifikasi 90%:10% dengan Indeks Twoing

```

=====
TEST SAMPLE CLASSIFICATION TABLE
=====

```

Actual Class	Predicted Class		Actual Total
	1	2	
1	137.00	49.00	186.00
2	81.00	162.00	243.00
PRED. TOT.	218.00	211.00	429.00
CORRECT	0.737	0.667	
SUCCESS IND.	0.303	0.100	
TOT. CORRECT	0.697		

```

      SENSITIVITY:          0.737      SPECIFICITY:          0.667
FALSE REFERENCE:          0.372  FALSE RESPONSE:          0.232
REFERENCE = "1", RESPONSE = "2"
-----

```

```

=====
TEST SAMPLE CLASSIFICATION PROBABILITY TABLE
=====

```

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.737	0.263	1.000
2	0.333	0.667	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	1322.00	326.00	1648.00
2	389.00	1824.00	2213.00
PRED. TOT.	1711.00	2150.00	3861.00
CORRECT	0.802	0.824	
SUCCESS IND.	0.375	0.251	
TOT. CORRECT	0.815		

SENSITIVITY: 0.802 SPECIFICITY: 0.824
 FALSE REFERENCE: 0.227 FALSE RESPONSE: 0.152
 REFERENCE = "1", RESPONSE = "2"

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.802	0.198	1.000
2	0.176	0.824	1.000

Output C9. Ketepatan Klasifikasi 95%:5% dengan Indeks Gini

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	63.00	24.00	87.00
2	43.00	84.00	127.00
PRED. TOT.	106.00	108.00	214.00
CORRECT	0.724	0.661	
SUCCESS IND.	0.318	0.068	
TOT. CORRECT	0.687		

SENSITIVITY: 0.724 SPECIFICITY: 0.661
 FALSE REFERENCE: 0.406 FALSE RESPONSE: 0.222
 REFERENCE = "1", RESPONSE = "2"

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.724	0.276	1.000
2	0.339	0.661	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	1229.00	518.00	1747.00
2	898.00	1431.00	2329.00
PRED. TOT.	2127.00	1949.00	4076.00
CORRECT	0.703	0.614	
SUCCESS IND.	0.275	0.043	
TOT. CORRECT	0.653		

SENSITIVITY: 0.703 SPECIFICITY: 0.614
 FALSE REFERENCE: 0.422 FALSE RESPONSE: 0.266
 REFERENCE = "1", RESPONSE = "2"

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	0.703	0.297	1.000
2	0.386	0.614	1.000

Output C10. Ketepatan Klasifikasi 95%:5% dengan Indeks Twoing

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	1	2	
1	63.00	24.00	87.00
2	43.00	84.00	127.00
PRED. TOT.	106.00	108.00	214.00
CORRECT	0.724	0.661	

SUCCESS IND. 0.318 0.068
 TOT. CORRECT 0.687

SENSITIVITY: 0.724 SPECIFICITY: 0.661
 FALSE REFERENCE: 0.406 FALSE RESPONSE: 0.222
 REFERENCE = "1", RESPONSE = "2"

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class 1	2	Actual Total
1	0.724	0.276	1.000
2	0.339	0.661	1.000

=====

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class 1	2	Actual Total
1	1229.00	518.00	1747.00
2	898.00	1431.00	2329.00
PRED. TOT.	2127.00	1949.00	4076.00
CORRECT	0.703	0.614	
SUCCESS IND.	0.275	0.043	
TOT. CORRECT	0.653		

=====

SENSITIVITY: 0.703 SPECIFICITY: 0.614
 FALSE REFERENCE: 0.422 FALSE RESPONSE: 0.266
 REFERENCE = "1", RESPONSE = "2"

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class 1	2	Actual Total
1	0.703	0.297	1.000
2	0.386	0.614	1.000

=====

Lampiran D. Pohon Klasifikasi Maksimal

Output D1. Informasi Pembentukan Pohon Klasifikasi dengan *Test Sample* Kombinasi Data Learning dan data Testing 95%:5%

Tree Number	Terminal Nodes	Test Set Relative Cost	Resubstitution Relative Cost	Complexity
1	512	0.671 ± 0.066	0.247	-1.000
2	507	0.671 ± 0.066	0.247	1.00E-005
3	504	0.663 ± 0.066	0.247	3.38E-005
4	500	0.663 ± 0.066	0.248	4.58E-005
5	479	0.640 ± 0.065	0.251	8.14E-005
6	461	0.648 ± 0.065	0.254	0.000105
7	439	0.648 ± 0.065	0.259	0.000117
8	433	0.648 ± 0.065	0.260	0.000129
9	402	0.648 ± 0.065	0.269	0.000153
10	382	0.636 ± 0.065	0.276	0.000163
11	376	0.636 ± 0.065	0.278	0.000177
12	364	0.636 ± 0.065	0.282	0.000189
13	354	0.636 ± 0.065	0.286	0.000201
14	317	0.636 ± 0.065	0.302	0.000225
15	314	0.636 ± 0.065	0.303	0.000249
16	298	0.636 ± 0.065	0.311	0.000260
17	275	0.648 ± 0.065	0.324	0.000296
18	270	0.648 ± 0.065	0.327	0.000311
19	263	0.648 ± 0.065	0.332	0.000325
20	257	0.651 ± 0.065	0.336	0.000356
21	237	0.651 ± 0.065	0.350	0.000368
22	224	0.656 ± 0.065	0.360	0.000381
23	215	0.640 ± 0.065	0.367	0.000392
24	207	0.663 ± 0.066	0.373	0.000403
25	197	0.663 ± 0.066	0.381	0.000415
26	172	0.655 ± 0.065	0.403	0.000439
27	168	0.655 ± 0.065	0.406	0.000457
28	153	0.643 ± 0.065	0.420	0.000470
29	149	0.643 ± 0.065	0.424	0.000493
30	141	0.643 ± 0.065	0.432	0.000511
31	115	0.651 ± 0.065	0.459	0.000530
32	113	0.651 ± 0.065	0.462	0.000547
33	107	0.667 ± 0.066	0.468	0.000582
34	96	0.667 ± 0.066	0.481	0.000597
35	94	0.675 ± 0.066	0.484	0.000618
36	86	0.656 ± 0.065	0.494	0.000654
37	78	0.651 ± 0.065	0.505	0.000697
38	74	0.651 ± 0.065	0.511	0.000707
39	72	0.651 ± 0.065	0.514	0.000725
40	70	0.666 ± 0.066	0.517	0.000762

Tree Number	Terminal Nodes	Test Set Relative Cost	Resubstitution Relative Cost	Complexity
41	62	0.659 ± 0.065	0.529	0.000779
42	58	0.647 ± 0.065	0.535	0.000797
43	56	0.647 ± 0.065	0.539	0.000869
44	52	0.655 ± 0.065	0.546	0.000905
45	48	0.643 ± 0.065	0.553	0.000940
46	46	0.651 ± 0.065	0.558	0.001
47	43	0.652 ± 0.065	0.564	0.001
48	41	0.652 ± 0.065	0.569	0.001
49	37	0.660 ± 0.066	0.578	0.001
50	36	0.660 ± 0.066	0.580	0.001
51	34	0.660 ± 0.066	0.586	0.001
52	32	0.668 ± 0.066	0.591	0.001
53	28	0.656 ± 0.065	0.603	0.002
54	27	0.648 ± 0.065	0.607	0.002
55	20	0.647 ± 0.065	0.636	0.002
56	19	0.625 ± 0.065	0.641	0.002
57	16	0.645 ± 0.065	0.658	0.003
58**	12	0.614 ± 0.064	0.682	0.003
59	11	0.630 ± 0.064	0.689	0.003
60	10	0.642 ± 0.064	0.696	0.004
61	9	0.627 ± 0.064	0.705	0.004
62	5	0.743 ± 0.066	0.749	0.005
63	3	0.806 ± 0.068	0.771	0.005
64	2	0.826 ± 0.068	0.805	0.017
65	1	1.000 ± 0.000	1.000	0.097

* Minimum Cost

** Optimal

Output D2. Informasi Pemangkasan Pohon Klasifikasi dengan *Test Sample* Kombinasi Data Learning dan data Testing 95%:5%

=====

TREE SEQUENCE

=====

Dependent variable: TERJANGK

Terminal Tree	Nodes	Test Set Relative Cost	Resubstitution Relative Cost	Complexity Parameter
1	512	0.671 +/- 0.066	0.247	0.000
56	19	0.625 +/- 0.065	0.641	0.002
57	16	0.645 +/- 0.065	0.658	0.003
58**	12	0.614 +/- 0.064	0.682	0.003
59	11	0.630 +/- 0.064	0.689	0.003
60	10	0.642 +/- 0.064	0.696	0.004
61	9	0.627 +/- 0.064	0.705	0.004
62	5	0.743 +/- 0.066	0.749	0.005
63	3	0.806 +/- 0.068	0.771	0.005

64	2	0.826 +/- 0.068	0.805	0.017
65	1	1.000 +/- 0.000	1.000	0.097

Initial misclassification cost = 0.500

Initial class assignment = 1

Output D3. Kesalahan Klasifikasi Data *Learning* dan Data *Testing*
 Pohon Klasifikasi Maksimal

Misclassification for Learn Data

Class	N Cases	N Mis-classed	Pct Error	Cost
2	1747	189	10.82	0.11
1	2329	321	13.78	0.14

Misclassification for Test Data

Class	N Cases	N Mis-classed	Pct Error	Cost
1	127	39	30.71	0.31
2	87	32	36.78	0.37

Lampiran E. Pohon Klasifikasi Optimal

Output E1. Kesalahan Klasifikasi Data *Learning* dan Data *Testing* Pohon Klasifikasi Optimal

Misclassification for Learn Data

Class	N Cases	N Mis-classed	Pct Error	Cost
2	1747	518	29.65	0.30
1	2329	898	38.56	0.39

Misclassification for Test Data

Class	N Cases	N Mis-classed	Pct Error	Cost
2	87	24	27.59	0.28
1	127	43	33.86	0.34

Output E2. Informasi Pemilahan Simpul Pohon Klasifikasi Optimal

=====

=====

=====

```

*****
*           Node 1: PEKERJAA           *
*           N: 4076                     *
*****

```

```

*****
*           Node 2           *           *           Node 6           *
*           N: 1845         *           *           N: 2231         *
*****

```

Node 1 was split on PEKERJAA

A case goes left if PEKERJAA = (1,2,3)

Improvement = 0.019

Complexity Threshold = 0.097

Node	Cases	Wgt Counts	Cost	Class
1	4076	4076.00	0.500	1
2	1845	1845.00	0.396	1
6	2231	2231.00	0.409	2

Class	Weighted Counts		
	Top	Left	Right
1	1747.00	985.00	762.00
2	2329.00	860.00	1469.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.500	0.604	0.409
2	0.500	0.396	0.591

Surrogate	Split	Assoc.	Improve.
1 PENDIDIK s 5,6,7		0.405	0.006
2 STATUS E s 2		0.258	.441287E-03
3 WILAYAH s 1		0.243	0.008
4 PENGOLAH s 2		0.144	0.002

```

5 UMUR_KK      s      48.965      0.139      0.001

Competitor      Split      Improve.
1 STATUS_K      2      0.016
2 JK_KK         1      0.011
3 PENDIDIK      5,7    0.008
4 WILAYAH       1      0.008
5 JUMLAH_P      115.000 0.007

```

```

*****
*           Node 2: STATUS_K          *
*           N: 1845                   *
*****

```

```

*****
*           Node 3                    *
*           N: 1731                   *
*****
=====
*           Terminal Node 5           *
*           N: 114                    *
=====

```

Node 2 was split on STATUS_K
A case goes left if STATUS_K = (2)
Improvement = 0.011 Complexity Threshold = 0.017

Node	Cases	Wgt	Counts	Cost	Class
2	1845		1845.00	0.396	1
3	1731		1731.00	0.370	1
-5	114		114.00	0.157	2

Class	Weighted Counts		
	Top	Left	Right
1	985.00	971.00	14.00
2	860.00	760.00	100.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.604	0.630	0.157
2	0.396	0.370	0.843

```

Surrogate      Split      Assoc.      Improve.
1 JK_KK        s 1      0.323      0.004

Competitor      Split      Improve.
1 PENGOLAH      1      0.006
2 JUMLAH_P      60.805 0.005
3 JK_KK         1      0.004
4 DAERAH_K      2      0.004
5 UMUR_KK       26.068 0.002

```

```

*****
*           Node 3: PENGOLAH          *
*           N: 1731                   *
*****

```

```

=====
*           Terminal Node 1           *
*           N: 1290                   *
=====
*****
*           Node 4                    *
*           N: 441                    *
*****

```

Node 3 was split on PENGOLAH
A case goes left if PENGOLAH = (1)
Improvement = 0.006 Complexity Threshold = 0.005

Node	Cases	Wgt	Counts	Cost	Class
3	1731		1731.00	0.370	1
-1	1290		1290.00	0.324	1
4	441		441.00	0.488	2

Weighted Counts			
Class	Top	Left	Right
1	971.00	787.00	184.00
2	760.00	503.00	257.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.630	0.676	0.488
2	0.370	0.324	0.512

Competitor	Split	Improve.
1 DAERAH_K	2	0.004
2 JUMLAH_P	55.000	0.004
3 PEKERJAA	1,3	0.002
4 SUMBER_A	2	0.002
5 PENDIDIK	1,2,3,5	0.002


```

*****
*           Node 4: DAERAH_K           *
*           N: 441                       *
*****

*****
*           Node 5                       *
*           N: 334                       *
*****

=====
*           Terminal Node 4             =
*           N: 107                       =
=====

```

Node 4 was split on DAERAH_K
A case goes left if DAERAH_K = (2)
Improvement = 0.003 Complexity Threshold = 0.008

Node	Cases	Wgt	Counts	Cost	Class
4	441		441.00	0.488	2
5	334		334.00	0.449	1
-4	107		107.00	0.278	2

Weighted Counts			
Class	Top	Left	Right
1	184.00	160.00	24.00
2	257.00	174.00	83.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.488	0.551	0.278
2	0.512	0.449	0.722

Surrogate	Split	Assoc.	Improve.
1 JUMLAH_P r	85.000	0.165	0.002

Competitor	Split	Improve.
1 JUMLAH_P	325.000	0.003
2 UMUR_KK	53.500	0.002
3 PENCEGAH	2	0.001
4 PENDIDIK	2,5,6,7	.957806E-03
5 PEKERJAA	1	.465433E-03

```

*****
*           Node 5: UMUR_KK           *
*           N: 334                     *
*****

```

```

=====
=      Terminal Node 2      =      Terminal Node 3      =
=      N: 290              =      N: 44                =
=====

```

Node 5 was split on UMUR_KK
 A case goes left if UMUR_KK <= 53.500
 Improvement = 0.002 Complexity Threshold = 0.004

Node	Cases	Wgt	Counts	Cost	Class
5	334		334.00	0.449	1
-2	290		290.00	0.412	1
-3	44		44.00	0.282	2

Class	Weighted Counts		
	Top	Left	Right
1	160.00	150.00	10.00
2	174.00	140.00	34.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.551	0.588	0.282
2	0.449	0.412	0.718

Competitor	Split	Improve.
1 JUMLAH_P	325.000	0.001
2 PENCEGAH	2	0.001
3 PENDIDIK	2,3,5,6,7	.400108E-03
4 PEKERJAA	1,3	.349639E-03
5 JK_KK	1	.321687E-03

```

*****
*           Node 6: STATUS_K           *
*           N: 2231                    *
*****

```

```

*****
*           Node 7           *
*           N: 1892          *
*****
=====
=      Terminal Node 12      =
=      N: 339                =
=====

```

Node 6 was split on STATUS_K
 A case goes left if STATUS_K = (2)
 Improvement = 0.004 Complexity Threshold = 0.005

Node	Cases	Wgt	Counts	Cost	Class
6	2231		2231.00	0.409	2
7	1892		1892.00	0.435	2
-12	339		339.00	0.258	2

Class	Weighted Counts		
	Top	Left	Right
1	762.00	692.00	70.00
2	1469.00	1200.00	269.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.409	0.435	0.258
2	0.591	0.565	0.742

Surrogate		Split	Assoc.	Improve.
1 JK_KK	s 1		0.385	0.004

Competitor		Split	Improve.
1 JK_KK	1		0.004
2 WILAYAH	1		0.003
3 LAYANAN	1		0.002
4 PENDIDIK	3,4,5		0.002
5 SUMBER_A	2		0.002

```

*****
*           Node 7: LAYANAN           *
*           N: 1892                    *
*****

```

```

*****
*           Node 8                     *
*           N: 1115                    *
*****
=====
*           Terminal Node 11           *
*           N: 777                     *
=====

```

Node 7 was split on LAYANAN

A case goes left if LAYANAN = (1)

Improvement = 0.002 Complexity Threshold = 0.007

Node	Cases	Wgt Counts	Cost	Class
7	1892	1892.00	0.435	2
8	1115	1115.00	0.477	2
-11	777	777.00	0.372	2

Class	Weighted Counts		
	Top	Left	Right
1	692.00	453.00	239.00
2	1200.00	662.00	538.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.435	0.477	0.372
2	0.565	0.523	0.628

Surrogate		Split	Assoc.	Improve.
1 PENGOLAH	s 1		0.048	.927340E-03
2 STATUS_E	s 1		0.033	.728466E-04

Competitor		Split	Improve.
1 WILAYAH	1		0.002
2 JUMLAH_P	209.500		0.002
3 SUMBER_A	2		0.002
4 PEKERJAA	4		0.002
5 JK_KK	1		0.002

```

*****
*           Node 8: JUMLAH_P           *
*           N: 1115                    *
*****

```

```

*****
*           Node 9                     *
*           N: 779                     *
*****
=====
*           Node 11                     *
*           N: 336                     *
=====

```

Node 8 was split on JUMLAH_P
 A case goes left if JUMLAH_P <= 214.500
 Improvement = 0.003 Complexity Threshold = 0.011

Node	Cases	Wgt Counts	Cost	Class
8	1115	1115.00	0.477	2
9	779	779.00	0.479	1
11	336	336.00	0.371	2

Class	Weighted Counts		
	Top	Left	Right
1	453.00	350.00	103.00
2	662.00	429.00	233.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.477	0.521	0.371
2	0.523	0.479	0.629

Competitor	Split	Improve.
1 JK_KK	1	0.002
2 PENDIDIK	3,4,5	0.002
3 WILAYAH	1	0.002
4 PEKERJAA	4,6,7	0.001
5 DAERAH_K	2	0.001

```

*****
*           Node 9: PENDIDIK           *
*           N: 779                     *
*****

```

```

*****
*           Node 10          *           =           Terminal Node 8           =
*           N: 526          *           =           N: 253           =
*****

```

Node 9 was split on PENDIDIK
 A case goes left if PENDIDIK = (3,4,5)
 Improvement = 0.003 Complexity Threshold = 0.014

Node	Cases	Wgt Counts	Cost	Class
9	779	779.00	0.479	1
10	526	526.00	0.417	1
-8	253	253.00	0.386	2

Class	Weighted Counts		
	Top	Left	Right
1	350.00	269.00	81.00
2	429.00	257.00	172.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.521	0.583	0.386
2	0.479	0.417	0.614

Surrogate	Split	Assoc.	Improve.
1 UMUR_KK s	66.500	0.135	.402123E-03

Competitor	Split	Improve.
1 JK_KK	1	0.003
2 JUMLAH_P	155.000	0.003
3 PEKERJAA	4,6,7	0.002
4 UMUR_KK	23.269	.940762E-03

5 DAERAH_K 2 .767390E-03

* Node 10: JK_KK *
* N: 526 *

```
=====
= Terminal Node 6 = = Terminal Node 7 =
= N: 506 = = N: 20 =
=====
```

Node 10 was split on JK_KK
A case goes left if JK_KK = (1)
Improvement = 0.002 Complexity Threshold = 0.003

Node	Cases	Wgt	Counts	Cost	Class
10	526		526.00	0.417	1
-6	506		506.00	0.402	1
-7	20		20.00	0.129	2

Weighted Counts			
Class	Top	Left	Right
1	269.00	267.00	2.00
2	257.00	239.00	18.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.583	0.598	0.129
2	0.417	0.402	0.871

Surrogate		Split	Assoc.	Improve.
1	JUMLAH_P r	25.000	0.048	.145935E-03

Competitor		Split	Improve.
1	JUMLAH_P	155.000	0.002
2	PEKERJAA	4,6,7	0.002
3	UMUR_KK	23.269	0.001
4	DAERAH_K	2	.611310E-03
5	SUMBER_A	2	.602776E-03

* Node 11: UMUR_KK *
* N: 336 *

```
=====
= Terminal Node 9 = = Terminal Node 10 =
= N: 295 = = N: 41 =
=====
```

Node 11 was split on UMUR_KK
A case goes left if UMUR_KK <= 65.179
Improvement = 0.002 Complexity Threshold = 0.004

Node	Cases	Wgt	Counts	Cost	Class
11	336		336.00	0.371	2
-9	295		295.00	0.324	2
-10	41		41.00	0.324	1

Weighted Counts			
Class	Top	Left	Right

1	103.00	78.00	25.00
2	233.00	217.00	16.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.371	0.324	0.676
2	0.629	0.676	0.324

Competitor	Split	Improve.
1 SUMBER_A	2	0.001
2 PEKERJAA	4, 6, 7, 8	.999667E-03
3 PENCEGAH	2	.661572E-03
4 WILAYAH	1	.604394E-03
5 DAERAH_K	2	.553251E-03

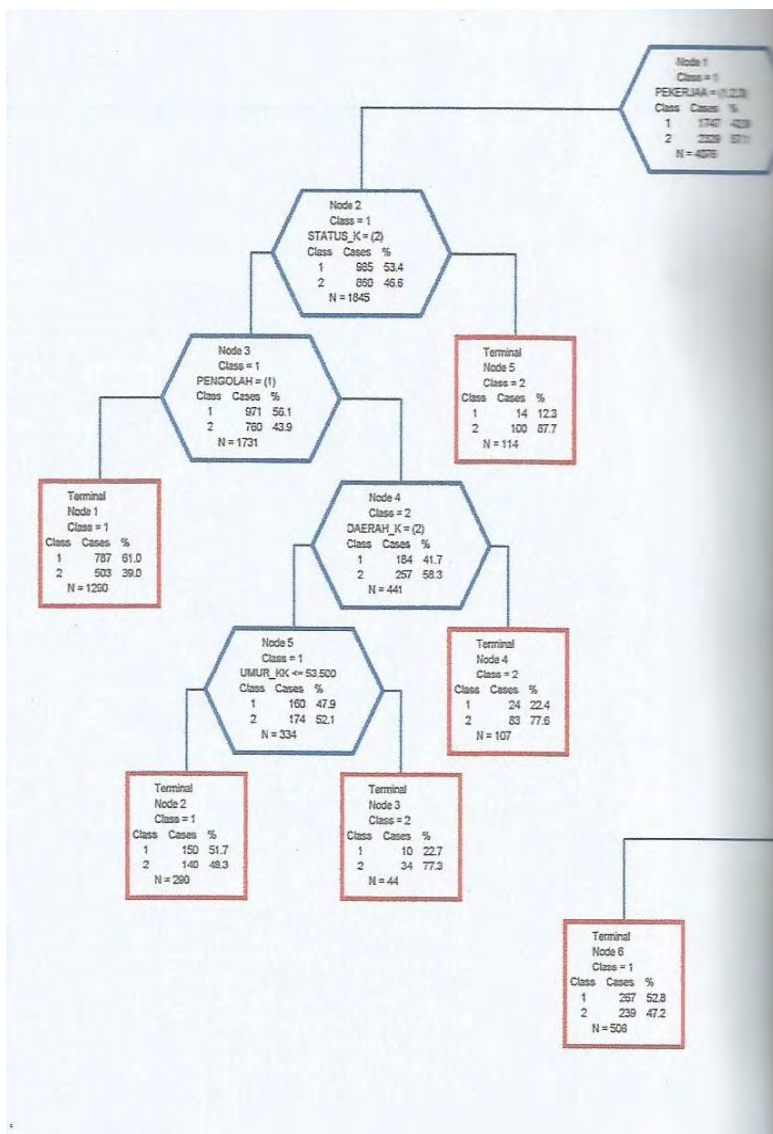
Output E3. Informasi Proporsi Setiap Kelas pada Setiap Simpul Pohon Klasifikasi Optimal

```
=====
TERMINAL NODE INFORMATION
=====
(Test Set)
```

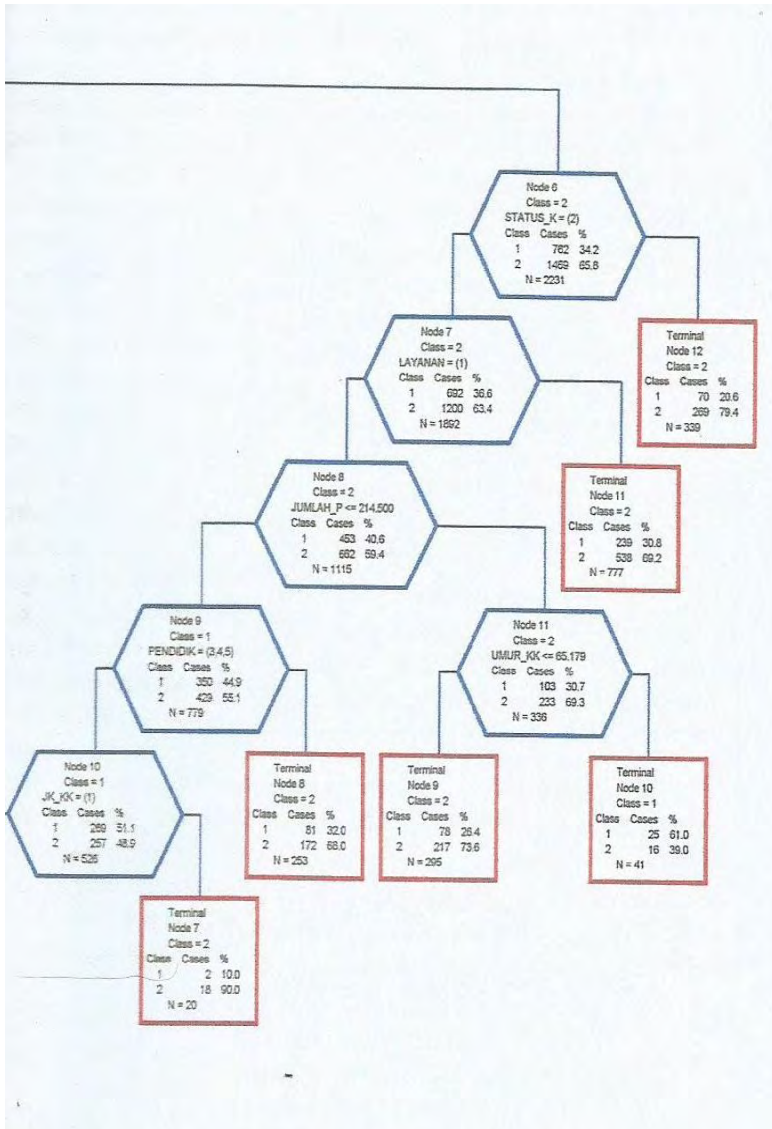
Node	Class	Wgt Count	N	Prob	Cost	Parent Complexity

1	1	1290.00 (57.00)	1290 57	0.333 0.290	0.324 (0.286)	0.005
	1	787.00 (36.00)	787 36	0.676 (0.714)		
	2	503.00 (21.00)	503 21	0.324 (0.286)		
2	1	290.00 (18.00)	290 18	0.073 0.084	0.412 (0.518)	0.004
	1	150.00 (7.00)	150 7	0.588 (0.482)		
	2	140.00 (11.00)	140 11	0.412 (0.518)		
3	2	44.00 (3.00)	44 3	0.010 0.015	0.282 (0.745)	0.004
	1	10.00 (2.00)	10 2	0.282 (0.745)		
	2	34.00 (1.00)	34 1	0.718 (0.255)		
4	2	107.00 (8.00)	107 8	0.025 0.031	0.278 (0.000)	0.008
	1	24.00 (0.00)	24 0	0.278 (0.000)		
	2	83.00 (8.00)	83 8	0.722 (1.000)		
5	2	114.00 (5.00)	114 5	0.025 0.021	0.157 (0.267)	0.017
	1	14.00 (1.00)	14 1	0.157 (0.267)		
	2	100.00 (4.00)	100 4	0.843 (0.733)		

6	1	506.00 (30.00)	506 30	0.128 0.153	0.402 0.284)	0.003
	1	267.00 (19.00)	267 19	0.598 0.716)		
	2	239.00 (11.00)	239 11	0.402 0.284)		
7	2	20.00 (2.00)	20 2	0.004 0.008	0.129 0.000)	0.003
	1	2.00 (0.00)	2 0	0.129 0.000)		
	2	18.00 (2.00)	18 2	0.871 1.000)		
8	2	253.00 (16.00)	253 16	0.060 0.067	0.386 0.173)	0.014
	1	81.00 (2.00)	81 2	0.386 0.173)		
	2	172.00 (14.00)	172 14	0.614 0.827)		
9	2	295.00 (23.00)	295 23	0.069 0.100	0.324 0.289)	0.004
	1	78.00 (5.00)	78 5	0.324 0.289)		
	2	217.00 (18.00)	217 18	0.676 0.711)		
10	1	41.00 (1.00)	41 1	0.011 0.006	0.324 0.000)	0.004
	1	25.00 (1.00)	25 1	0.676 1.000)		
	2	16.00 (0.00)	16 0	0.324 0.000)		
11	2	777.00 (37.00)	777 37	0.184 0.164	0.372 0.351)	0.007
	1	239.00 (10.00)	239 10	0.372 0.351)		
	2	538.00 (27.00)	538 27	0.628 0.649)		
12	2	339.00 (14.00)	339 14	0.078 0.062	0.258 0.369)	0.005
	1	70.00 (4.00)	70 4	0.258 0.369)		
	2	269.00 (10.00)	269 10	0.742 0.631)		

Output E4. Struktur Pohon Klasifikasi Optimal

Output E5. Struktur Pohon Klasifikasi Optimal (Lanjutan)



Output E6. Penjelasan Pohon Klasifikasi Optimal

Variabel pekerjaan kepala keluarga (X_6) memilah simpul utama (simpul 1) menjadi simpul kiri dan simpul kanan dengan ketentuan pekerjaan kepala keluarga yaitu PNS/TNI/POLRI-/BUMD, pegawai swasta dan wiraswasta (kategori 1, 2 dan 3) akan dipilah menjadi simpul kiri (simpul 2), sedangkan jika pekerjaan kepala keluarga sebagai petani, nelayan, buruh, lainnya dan tidak bekerja (kategori 4, 5, 6, 7 dan 8), akan dipilah menjadi simpul kanan (simpul 6). Diperoleh hasil bahwa ada sebanyak 1845 rumah tangga yang kepala keluarganya bekerja sebagai PNS/TNI/POLRI/BUMD, pegawai swasta dan wiraswasta menjadi anggota simpul kiri (simpul 2) dan sisanya sebanyak 2231 rumah tangga yang kepala keluarganya bekerja sebagai petani, nelayan, buruh, lainnya dan tidak bekerja yang menjadi anggota simpul kanan (simpul 6).

Simpul 2 yang beranggotakan 1845 rumah tangga dengan kepala keluarga yang bekerja sebagai PNS/TNI/POLRI/BUMD, pegawai swasta dan wiraswasta, selanjutnya dipilah menjadi simpul baru kiri dan kanan menurut status kawin kepala keluarga (X_3). Jika status kawin kepala keluarganya menikah, maka akan dipilah menjadi anggota simpul kiri baru (simpul 3). Namun jika status kawin kepala keluarga belum menikah, hidup bersama, cerai hidup, hidup terpisah, dan cerai mati, maka rumah tangga tersebut akan dipilah menjadi simpul kanan baru (simpul terminal 5). Diantara 1845 rumah tangga anggota simpul 2, diperoleh hasil ada sebanyak 1731 rumah tangga menjadi anggota simpul 3 dengan karakteristik kepala keluarga yang bekerja sebagai PNS/TNI/POLRI/BUMD, pegawai swasta dan wiraswasta serta status kawin kepala keluarganya adalah menikah. Sisanya sebanyak 114 rumah tangga yang menjadi anggota simpul terminal 5 dengan karakteristik kepala keluarga yang bekerja sebagai PNS/TNI/POLRI/BUMD, pegawai swasta dan wiraswasta serta status kawin kepala keluarganya adalah belum menikah, hidup bersama, cerai hidup, hidup terpisah, dan cerai mati.

Simpul 3 yang beranggotakan 1731 rumah tangga dengan status kawin kepala keluarganya adalah menikah, selanjutnya dipilah menjadi simpul baru kiri dan kanan menurut kebiasaan mengolah air minum sebelum dikonsumsi. Jika rumah tangga melakukan pengolahan air minum sebelum dikonsumsi, maka rumah tangga tersebut akan dipilah ke simpul baru kiri (simpul terminal 1). Sedangkan jika rumah tangga tidak melakukan pengolahan air minum sebelum dikonsumsi, maka akan dipilah ke simpul kanan baru (simpul 4). Diperoleh hasil bahwa diantara

1731 rumah tangga anggota simpul 3, terdapat 1290 rumah tangga yang menjadi anggota simpul terminal 1 dengan karakteristik kepala keluarga yang bekerja sebagai PNS/TNI-/POLRI/BUMD, pegawai swasta dan wiraswasta, status kawin kepala keluarganya adalah menikah dan rumah tangga melakukan pengolahan air minum sebelum dikonsumsi. Sedangkan sisanya sebanyak 441 rumah tangga dipilah ke simpul 4. Begitu seterusnya hingga penjelasan pada simpul 11 yang dipilah menjadi simpul terminal 9 dan simpul terminal 10.

Lampiran F. Surat Keterangan Data



**KEMENTERIAN KESEHATAN RI
BADAN PENELITIAN DAN
PENGEMBANGAN KESEHATAN**

Jalan Percetakan Negara No. 29 Jakarta 10560 Kotak Pos 1226

Telepon: (021) 4261088 Faksimile: (021) 4243933

E-mail: sesben@litbang.depkes.go.id, Website: <http://www.litbang.depkes.go.id>

SURAT PERNYATAAN

Pada hari ini Rabu tanggal 22 bulan April tahun 2015 yang bertanda tangan di bawah ini:

Nama	: Ayu Widya Ningrum
Alamat E-mail	: ayuwidyaningrum121@gmail.com
Hp	: 089676963408
NIP/ NPM	: 1311 100 121
Pekerjaan	: Mahasiswa
Instansi	: Institut Teknologi Sepuluh Nopember
Alamat Kantor	: Kampus ITS Sukolilo Surabaya, 60111
Universitas (mahasiswa)	: Institut Teknologi Sepuluh Nopember
Judul Penelitian	: Classification And Regression Tree untuk Pengklasifikasi Penderita Penyakit Malaria di Provinsi Papua Barat dengan Pra-Pemrosesan Synthetic Minority Oversampling Technique

Menyatakan dengan sesungguhnya bahwa :

1. Saya sanggup dan bersedia untuk mematuhi ketentuan-ketentuan yang telah ditetapkan dalam melakukan kegiatan penelitian dan pengembangan sesuai dengan Undang-Undang Kesehatan Nomor 36 Tahun 2009 tentang Kesehatan dan Peraturan Pemerintah Nomor 39 Tahun 1995 tentang Penelitian dan Pengembangan Kesehatan
2. Saya telah menerima Subset Data hasil kegiatan penelitian Riset Kesehatan Dasar 2013 yang direkam dalam media elektronik yang dibuat oleh Badan Penelitian dan Pengembangan Kesehatan;
3. Data hasil penelitian yang saya peroleh akan saya gunakan untuk kepentingan Tugas Akhir/ Skripsi sehingga saya:
 - a. tidak akan membuat salinan dari data tersebut untuk keperluan lain dan pihak lain atau mengalihkan data tersebut pada pihak lain;
 - b. akan mempergunakan data tersebut hanya untuk 1 (satu) topik judul penelitian, sesuai dengan persetujuan yang diberikan secara formal oleh Badan Litbang Kesehatan;
 - c. apabila saya menggunakan data untuk keperluan lain selain dari ketentuan di atas, harus mengajukan kembali secara formal kepada Kepala Badan Litbang Kesehatan;
 - d. akan melakukan komunikasi dengan pihak Laboratorium Manajemen Data untuk pemahaman variabel subset data;
 - e. untuk melakukan publikasi hasil analisis, saya sanggup dan bersedia untuk terlebih dahulu memperhatikan etika dan manfaat bagi kepentingan masyarakat
4. Saya berkewajiban untuk menyerahkan hasil analisis kepada Laboratorium Manajemen Data Badan Penelitian dan Pengembangan Kesehatan.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa adanya unsur paksaan dari pihak manapun dan apabila dikemudian hari ternyata terjadi penyimpangan dari pernyataan saya tersebut, maka hak penggunaan data dan publikasi dinyatakan batal demi hukum.

Mengetahui,
Ketua/Wakil Lab. Mandat

[Signature]
Ika Dharmayanti

Pembuat Set Data

[Signature]
Olvin Haingol

Penerima Data

METERAI
TEMPEL
3
C8778ACF5066600
6000
DJP
(AYU WIDYA NINGRUM)

